Do AI-Algorithmic Traders Lead to Market Instability? A Multi-Agent Reinforcement Learning Approach

Yang Fan, Markus Pelger, Xintong Yu

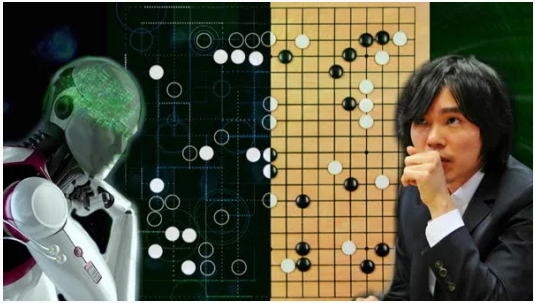
Stanford University, Stanford University and NBER, Columbia University

AI-Algorithmic Trading

AI-Algorithmic Trading: reinforcement learning (RL) algorithm + algorithmic trading

AlphaGo:

- Deep learning reinforcement learning algorithm (DL-RL)
  = computational agent learns optimal decisions for complex problems
- AlphaGo outperforms best human Go player
- Number of possible moves in Go $\approx 10^{170} \Rightarrow$ exceeds humans' cognitive capacity



This Paper

Algorithmic trading is widely adopted:

- 70-80 percent of US equity market volume in the last 5 to 10 years.
- Market participants increasingly sophisticated and open to adopt AI based systems

Research question: What happens when AI algorithms trade against each other?

This paper:

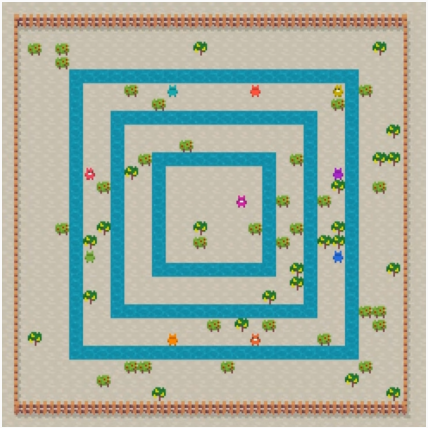
1. Conceptual framework to simulate equilibrium prices for AI algorithmic trading
2. Novel deep learning reinforcement learning algorithm (DL-RL) for trading

Findings:

1. AI algorithmic trading can lead to inefficient and unstable market prices
2. Informed AI-traders manipulate less informed AI-traders leading to local price bubbles
3. This new risk is more severe if information advantage and pricing power of informed AI-traders is large and if only few uninformed AI-traders are in the market

Motivation for Simulation Approach - DeepMind's Fruit Market

- **Analytical solution** for equilibrium with AI agents generally infeasible
- **Simulations** for multi-agent RL provides provides understanding of equilibrium
- **Motivating example**: "Emergent Bartering Behaviour in Multi-Agent Reinforcement Learning" (*M. Johanson et al. (2022)*)
- **Fruit Market**: deep RL agents learn how to produce, trade, and consume resources in order to maximize their individual reward.
- Some agents learn to arbitrage and specialize in the transportation of fruits



Some Related Work

Algorithmic collusion:

- "AI-Powered Trading, Algorithmic Collusion, and Price Efficiency" (*Dou, Goldstein and Ji (2023)*)
- "Artificial Intelligence, Algorithmic Pricing, and Collusion" (*Calvano et al. (2022)*)

⇒ Based on Q-Learning with simple discrete actions and states

⇒ "Collusion of mistakes" is corner case for simple homogeneous AI traders

⇒ Novel risk of AI manipulation requires "smarter" (DL-RL) and heterogeneous AI's

Pricing bubbles:

- Large literature including *Harrison and Kreps (1979), Blanchard (1979), Camerer (1989), Abreu and Brunnermeier (2003), Jarrow (2015)*

⇒ Model the emergence of bubbles through rule-based agents

⇒ We show that AI learns to generate price bubbles

Mean-field game theory:

- Emerging literature including *Cartea, Chang and Penalva (2022), Cartea et al (2022)*
- Analytical collusion results for stylized models with many small and homogeneous agents

⇒ We study more complex AI agents with market power

⇒ We identify novel risk different from collusion

Model

Model Setup

Market:

- One asset with **fundamental value** ν and **market price** p
- Periodic events that reveal fundamental value ν - similar to earnings events
- At revelation time market price $p = \nu$ to incorporate public information.

Market participants

1. **Informed AI traders**, who observe the true fundamental value ν
2. **Uninformed AI traders**, who learn about ν by observing prices $\{p_0, \ldots, p_{t-1}\}$, the net market orders $\{z_0, \ldots, z_{t-1}\}$, and noisy signal about ν.
3. **Market makers**, who determine the price following a linear equilibrium model consistent with the *Kyle (1985) model*:

$$p_t = p_{t-1} + \lambda z_t.$$

- Reduced form model to focus on novel model components
- Here, we fix λ, to reflect the liquidity/market impact of trading
- This pricing mechanism penalizes frivolous actions from agents.

AI Trading Agents

Actions:

- Continuous action $a_i \in [-1, 1]$, which is then mapped to $\{-k_i, 0, k_i\}$, where k_i is the buying power of the agent. Positions can accumulate.
- We could easily generalize it to $\{-k_i, -k_i + 1, \ldots, k_i - 1, k_i\}$ without increasing model size.

Information Set:

- All AI traders observe price, volume, technical indicators, balance, position, time within the trading periods
- **Informed AI** observes true fundamental value $\nu_{obs} = \nu$
- **Uninformed AI** observes noisy fundamental value: $\nu_{obs} = \nu + \epsilon$ for different noise levels ϵ

Rewards:

- At the end of each revelation episode we force liquidation at price ν and compute the final balance of each each agent
- Agents maximize cumulative change in cash balance = terminal reward
- Difficult to train with only terminal reward - equivalent reformulation of the problem with partial rewards at non-terminal steps which allows effective training of RL models

$\Rightarrow$ Correct formulation of reward is a crucial and non-trivial problem!

Deep Learning RL

Deep RL:

- Each AI agent has independent policy estimated by deep neural nets.
- Each policy is individually optimized by Proximal Policy Optimization (PPO)
- Difficult to optimize due to non-stationarity in our non-cooperative environment.

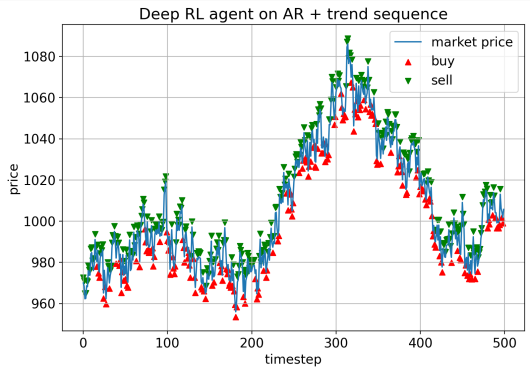
Training and Evaluation:

- Tuning parameters (learning rate and entropy coefficients) are fine-tuned by letting AI play against stochastic rule-based traders
- AI traders face each other for given tuning parameters and to learn optimal equilibrium policies
- Evaluate the converged models in the final game

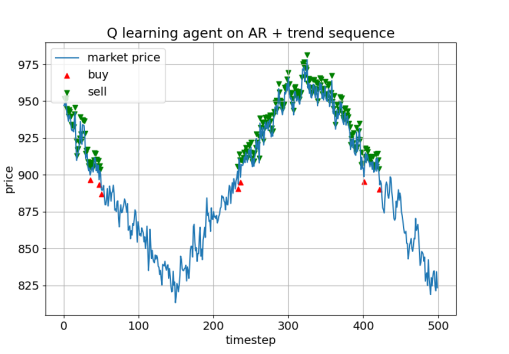
Why not Q-Learning:

- Q-learning uses discrete state space which leads to exploding dimensionality
- Our setup has 20 continuous variables: price, volumes, balance, time till until revelation, position, and technical indicators as continuous variables. Discretizing can lead to $2^{20} \sim 1$ million states. (*Dou et al (2023)* require up to 50 billion steps for training much simpler model)
- Q-learning not designed to learn complex price patterns and behavior

Illustration of AI Models



Deep RL

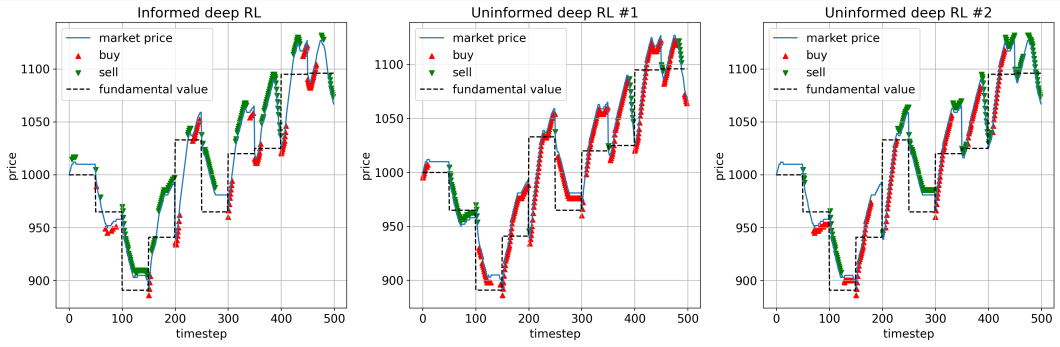


Q-Learning

- Deep RL takes advantage of fundamental price deviations and local patterns
- Note that, exploiting trend patterns requires multi-step planning
- Simple Q-learning not designed to learn complex price trend based patterns

Main Findings

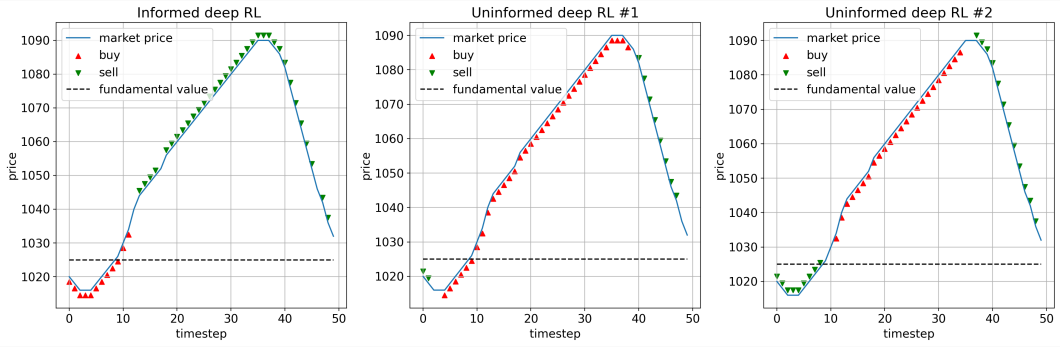
Manipulative Behavior from Informed AI



Equilibrium trading behavior for 1 informed and 2 uninformed AI traders:

- $\Rightarrow$ The informed AI buys even when $\nu < p$
- $\Rightarrow$ Two uniformed AIs create self-perpetuating upward trend
- $\Rightarrow$ higher profit for informed AI through manipulation

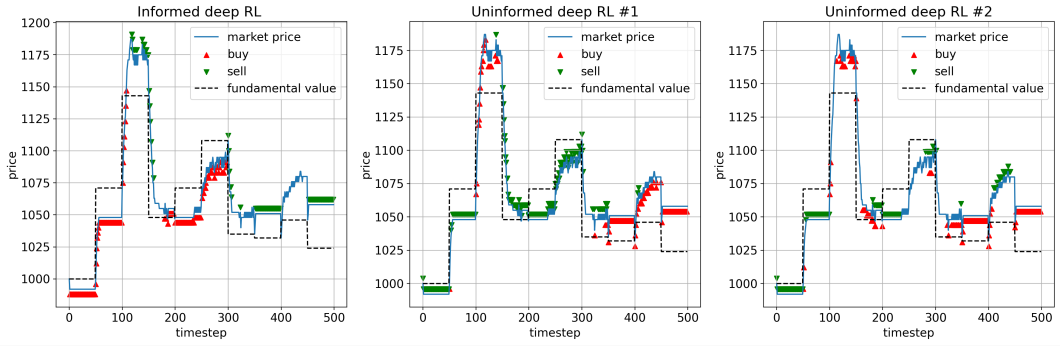
Manipulative Behavior from Informed AI (zoomed in)



Equilibrium trading behavior for 1 informed and 2 uninformed AI traders:

- $\Rightarrow \Rightarrow$ The informed AI buys even when $\nu < p$
- $\Rightarrow$ Two uniformed AIs create self-perpetuating upward trend
- $\Rightarrow$ higher profit for informed AI through manipulation

Strategic Lack of Price Correction

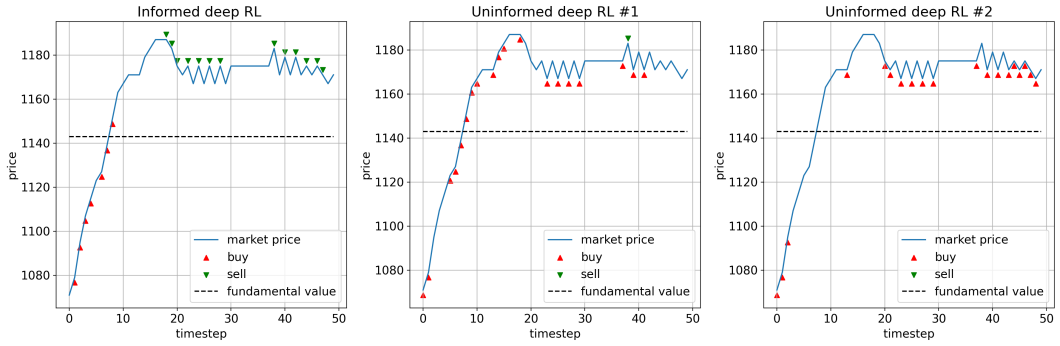


Equilibrium trading behavior for 1 informed and 2 uninformed AI traders:

⇒ Informed AI does not immediately sell when $\nu < p$, waiting for larger price gap to occur.

⇒ Inefficient equilibrium price through strategic price deviations

Strategic Lack of Price Correction (zoomed in)

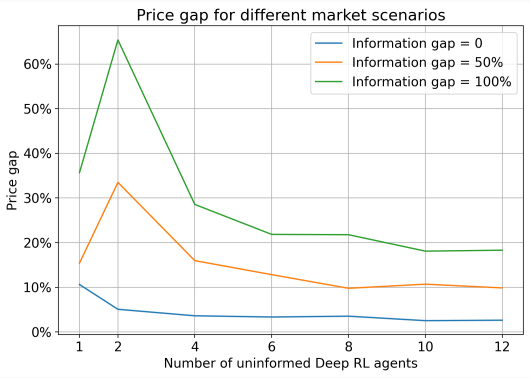


Equilibrium trading behavior for 1 informed and 2 uninformed AI traders:

$\Rightarrow$ Informed AI does not immediately sell when $\nu < p$, waiting for larger price gap to occur.

$\Rightarrow$ Inefficient equilibrium price through strategic price deviations

Market Inefficiency for Different Scenarios

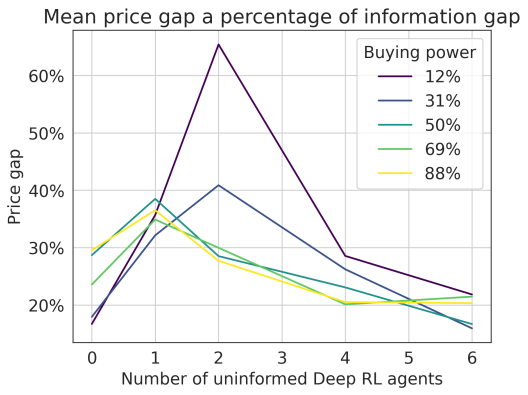


Average price gap between market price and fundamental value:

- 1 informed AI and n uninformed AI's with different noise levels ϵ in fundamental value
- Largest price gap for two uninformed AI's $\Rightarrow$ easiest to incur self-perpetuating trend
- Largest price gap for large information gap (large ϵ) $\Rightarrow$ easier to manipulate uninformed AI's

$\Rightarrow$ More efficiency for homogeneous information and many traders

Varying the Buying Power of Informed AI Trader



Average price gap between market price and fundamental value:

- Informed AI 1/8 of buying power; remaining buying power evenly split among uninformed AI's.
- Here we gradually increase the buying power of the informed AI (information gap is set to 50%)

$\Rightarrow$ Larger price inefficiencies when informed AI has smaller market share

$\Rightarrow$ Need sufficiently large but small number of uninformed AI's for manipulation

Conclusion

Conclusion

Study of "smart" AI algorithmic trading:

- Conceptual framework to simulate AI trading
- "Smart" deep learning RL traders with heterogeneous information and buying power

Market Inefficiencies induced by informed AI traders:

- Active manipulation and strategic lack of price correction by informed AI traders
- Largest inefficiencies for large information gap and few sufficiently large uninformed AI traders

Policy Implications:

- Promote regulatory oversight through sandbox simulation
- Promote competition and sufficiently many not too concentrated algorithmic traders
- Access of accurate information about fundamental value for all traders
- Awareness of unintended consequences from black box algorithms as they get smarter