

# Reinforcement Learning in Finance and Automated Trading

**Rahul Savani**

**Alan Turing Institute / University of Liverpool**

**ML x Finance @ OxML, 10th July 2023**

# Reinforcement Learning for Trading in Limit Order Book Environments

**Rahul Savani**

**Alan Turing Institute / University of Liverpool**

**ML x Finance @ OxML, 10th July 2023**

# Reinforcement Learning for Trading in Limit Order Book Environments

**Rahul Savani**

**Alan Turing Institute / University of Liverpool**

Parts based on joint work with:

Andrea Coletta, **Martin Herdegen**, **Joe Jerome**, John Fearnley,  
Andreas Koukorinis, **Leandro Sánchez-Betancourt**, Tom Spooner,  
Svitlana Vytrenko

Funding acknowledgements: EPSRC, J.P.Morgan AI Research

**ML x Finance @ OxML, 10th July 2023**

Financial applications of Reinforcement Learning (RL)

Introduction to Limit Order Books

Types of Limit Order Book RL environment

- Market replay

- Conditional generative models

- Mathematical models

Mathematical models of limit order book trading

Tutorial on mbt-gym (model-based-trading gym)

## Financial applications of Reinforcement Learning (RL)

Introduction to Limit Order Books

Types of Limit Order Book RL environment

- Market replay

- Conditional generative models

- Mathematical models

Mathematical models of limit order book trading

Tutorial on mbt-gym (model-based-trading gym)

# Applications of Reinforcement Learning in Finance

- ▶ Portfolio Allocation / Optimization
- ▶ Hedging (and Option Pricing)
- ▶ Optimal Execution (and Smart Order Routing)
- ▶ Market Making

**Recent advances in reinforcement learning in finance**

**Ben Hambly, Renyuan Xu, Huining Yang**

**Mathematical Finance (2023)** – 67 pages

# Our focus – limit order books

- ▶ To train good policies with RL **lots of experience is needed**  
(**sample inefficiency**)
- ▶ **Limit order books** (LOBs) are a **widespread market mechanism** relevant for **data-rich high frequency finance**
- ▶ **Intro to LOBs**: First, we will look at the limit order books as a market mechanism in detail
- ▶ In this context, we will also explain the role and challenge faced by a market maker
- ▶ **Ways to model LOBs**: Then, we will then discuss different ways to create **LOB environments** for RL
- ▶ **Hands-on RL demo**: Finally, focusing on one of these ways, I will give an RL demo using **mbt\_gym** (freely available)



Financial applications of Reinforcement Learning (RL)

## Introduction to Limit Order Books

Types of Limit Order Book RL environment

- Market replay

- Conditional generative models

- Mathematical models

Mathematical models of limit order book trading

Tutorial on mbt-gym (model-based-trading gym)

# Order driven, aka limit order book markets

- ▶ In **order-driven markets** (as opposed to quote-driven markets):  
**all** buyers and sellers display prices at which they wish to buy or sell a particular security, as well as the amounts of the security desired to be bought or sold
- ▶ **Buy** orders are called **bids**; **sell** orders are called **asks** or **offers**
- ▶ In the academic literature limit order books markets are sometimes called **continuous double auctions**
- ▶ **This market mechanism is extremely widespread**; it is used in, e.g., stock, derivatives, and sports betting markets

# Limit and market orders

Two orders types:

**market** and **limit** orders, both specified by **direction** (buy/sell) & **volume**

Additionally a **limit order** has a **limit price**

Unmatched limit orders sit in the **limit order book**

	Price	Asks
	101.50	5
	101.25	11
	101.00	12
	100.50	
	100.25	22
	100.00	
14	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

# Best prices, midprice, and spread

	Price	Asks
	101.50	5
	101.25	11
	101.00	12
	100.50	
Best ask	100.25	22
	100.00	
14	99.75	
30	99.50	Best bid
3	99.25	
11	99.00	
Bids		

The **spread** is the difference between the best ask and best bid:

$$100.25 - 99.75 = 0.5$$

The **midprice** is the average of the best ask and best bid:

$$0.5 \cdot (100.25 + 99.75) = 100$$

# Market makers provide liquidity

	Price	Asks
	101.50	5
	101.25	11
MM ask	101.00	12
	100.50	
Best ask	100.25	22
	100.00	
14	99.75	MM bid
30	99.50	
3	99.25	
11	99.00	
Bids		

A **market maker provides liquidity** by placing bids and asks  
(some or all of the 12 and 14 units of volume in the example)

# How do they make money?

	Price	Asks
	101.50	5
	101.25	11
MM ask	101.00	12
	100.50	
Best ask	100.25	22
	100.00	
14	99.75	MM bid
30	99.50	
3	99.25	
11	99.00	
Bids		

A **market maker** tries to repeatedly “**make the spread**”

# How do they make money?

	Price	Asks
	101.50	5
	101.25	11
MM ask	101.00	12
	100.50	
Best ask	100.25	22
	100.00	
14	99.75	MM bid
30	99.50	
3	99.25	
11	99.00	
Bids		

requires others to **“cross the spread”** in **both** directions, i.e.,  
someone lifts MM bid; someone **else** lifts MM ask

# Market order example

	Price	Asks
	101.50	5
	101.25	11
	101.00	12
	100.50	
	100.25	22
	100.00	
18	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

**Market order to buy 40**

$$\frac{1}{40} (100.25 \times 22 + 101.00 \times 12 + 101.25 \times 6) = 100.625$$



# Spread has widened

	Price	Asks
	101.50	5
	101.25	5
	101.00	
	100.50	
	100.25	
	100.00	
	99.75	
	99.50	
	99.25	
	99.00	
14		
30		
3		
11		
Bids		

last MM sale

old MM bid

**What happens next?** The spread will narrow typically, but how?

The MM would like the asks to be **replenished**...

# Midprice goes up

	Price	Asks
	101.50	5
	101.25	5
4	101.00	
10	100.50	
	100.25	
7	100.00	
14	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

last MM sale

old MM bid

...but instead **bids may fill the spread**

# Toxic order flow and inventory risk

	Price	Asks
	101.50	5
	101.25	5
4	101.00	
10	100.50	
	100.25	
7	100.00	
14	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

last MM sale

old MM bid

**Inventory risk:** may amass undesired position in trending markets

# Toxic order flow and inventory risk

	Price	Asks
	101.50	5
	101.25	5
4	101.00	
10	100.50	
	100.25	
7	100.00	
14	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

last MM sale

old MM bid

**Toxic order flow:** adverse selection of trading against better-informed counterparties

# Developing an automated market making strategy

	Price	Asks
	101.50	5
	101.25	5
4	101.00	
10	100.50	
	100.25	
7	100.00	
14	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

To effectively develop and test a market making strategy, one would like to be able to simulate **how the market reacts** to the actions of the market maker

# Developing an automated market making strategy

	Price	Asks
	101.50	5
	101.25	5
4	101.00	
10	100.50	
	100.25	
7	100.00	
14	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

Reinforcement Learning from scratch through live trading is generally infeasible; some form of **simulated environment** is used

# Developing an automated market making strategy

	Price	Asks
	101.50	5
	101.25	5
4	101.00	
10	100.50	
	100.25	
7	100.00	
14	99.75	
30	99.50	
3	99.25	
11	99.00	
Bids		

**Reinforcement Learning** is able to effectively **exploit weaknesses** in the **environment** (simulator)

Financial applications of Reinforcement Learning (RL)

Introduction to Limit Order Books

Types of Limit Order Book RL environment

- Market replay

- Conditional generative models

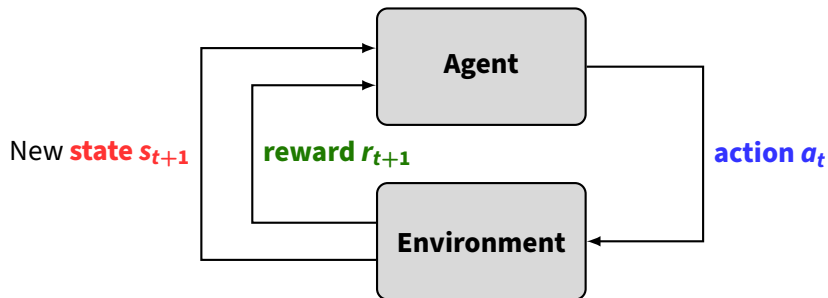
- Mathematical models

Mathematical models of limit order book trading

Tutorial on mbt-gym (model-based-trading gym)



# Reinforcement learning environment



How should an **agent** take **actions** in an **environment** so as to maximize some notion of **reward**?

## LOB environments: Three main possibilities

1. Backtesting – market replay
2. Conditional generative models
3. Mathematical models
4. (Agent-based models)

# Data-replay based LOB simulator (“backtesting”)

- ▶ **Replay historic** limit order book **data**
- ▶ Simulate a **queue at each level of the order book**
- ▶ Allow backtested agent to place orders in the book
- ▶ **Process transactions and cancellations** from both fixed data **and** the backtested agent
- ▶ Obviously difficulty: **the replayed data does not react to the actions of our backtested agent**

# A second issue: cancellations and queue modelling

## Market Making via Reinforcement Learning

AAMAS 2018

Thomas Spooner, John Fearnley, Rahul Savani, Andreas  
Koukorinis

assumes that cancellations happen uniformly across queue  
but, evidence shows they tend to happen near back of queue

## A second issue: cancellations and queue modelling

### **Market Making via Reinforcement Learning**

**AAMAS 2018**

**Thomas Spooner, John Fearnley, Rahul Savani, Andreas Koukorinis**

assumes that cancellations happen uniformly across queue

but, evidence shows they tend to happen near back of queue

### **Market Making with Scaled Beta Policies**

**ICAIF 2022**

**Joseph Jerome, Gregory Palmer, Rahul Savani**

highly realistic simulator using LOBSTER data (but still no reactivity)

## A second issue: cancellations and queue modelling

### **Market Making via Reinforcement Learning**

**AAMAS 2018**

**Thomas Spooner, John Fearnley, Rahul Savani, Andreas Koukorinis**

assumes that cancellations happen uniformly across queue

but, evidence shows they tend to happen near back of queue

### **Market Making with Scaled Beta Policies**

**ICAIF 2022**

**Joseph Jerome, Gregory Palmer, Rahul Savani**

highly realistic simulator using LOBSTER data (but still no reactivity)

Ignoring the reactivity issue, many papers that build market making agents do not simulate queues fully realistically

# rl4mm repo (requires LOBSTER data)

The screenshot shows the GitHub repository page for JJJerome/rl4mm. At the top, the repository name is displayed with a search bar and navigation icons. Below this, there are tabs for Issues, Pull requests, Actions, Projects, Security, and Insights. The repository is public and has 3 watchers, 2 forks, and 19 stars. The main branch is 'main' with 1 branch and 0 tags. A commit history table is shown with columns for file changes, commit message, and time. The 'About' section on the right describes the repository's purpose and includes links to the README and license.

JJJerome / rl4mm

Search Type to search

Issues Pull requests Actions Projects Security Insights

rl4mm Public

Watch 3 Fork 2 Starred 19

main 1 branch 0 tags

Go to file Add file Code

JJJerome Add CODE_OF_CONDUCT.md and CONTRIBUTING.md			384b318 on Nov 18, 2022	4 commits
checks	Initial public commit			10 months ago
docker	Initial public commit			10 months ago
examples	Add note about LOBSTER data to notebook			8 months ago

**About**

A repository for simulating limit order book dynamics from historical data and using it to train a reinforcement learning agent to make markets.

Readme

BSD-3-Clause license

<https://github.com/JJJerome/rl4mm>

# Conditional Generative Model as the simulator

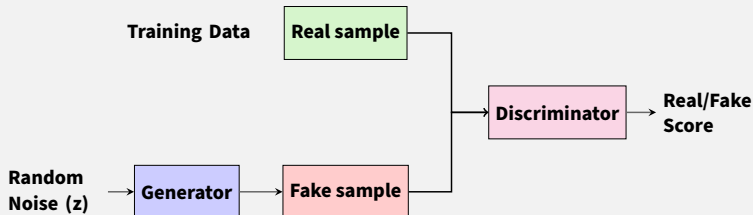
- ▶ Simulator based on **conditional generative model** that was trained on real order book data
- ▶ **Promises both realism and reactivity**



# Conditional Generative Model as the simulator

- ▶ Simulator based on **conditional generative model** that was trained on real order book data
- ▶ **Promises both realism and reactivity**

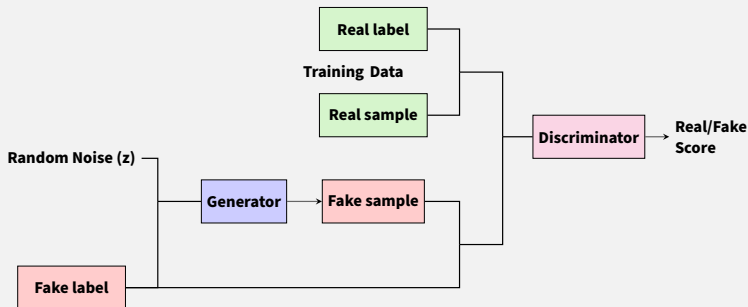
## (unconditional) Generative Adversarial Network (GAN)



# Conditional Generative Model as the simulator

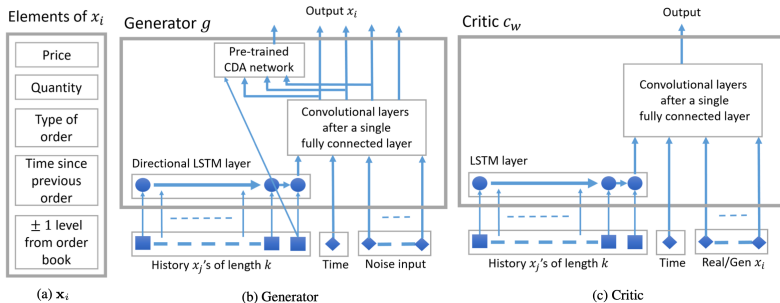
- ▶ Simulator based on **conditional generative model** that was trained on real order book data
- ▶ **Promises both realism and reactivity**

## Conditional Generative Adversarial Network (CGAN)



# Example setup

## Generating Realistic Stock Market Order Streams [Li et al, AAAI'20]



- ▶ Primarily tested by realism of outputs **in isolation**
- ▶ For building a trading agent using the CGAN we want: **interactive realism**

# More work to be done

## **Conditional Generators for Limit Order Book Environments: Explainability, Challenges, and Robustness**

**Andrea Coletta, Joseph Jerome, Rahul Savani, Svitlana Vyetenko**

CoRR abs/2306.12806 (2023)

# More work to be done

## Conditional Generators for Limit Order Book Environments: Explainability, Challenges, and Robustness

Andrea Coletta, Joseph Jerome, Rahul Savani, Svitlana Vyetenko

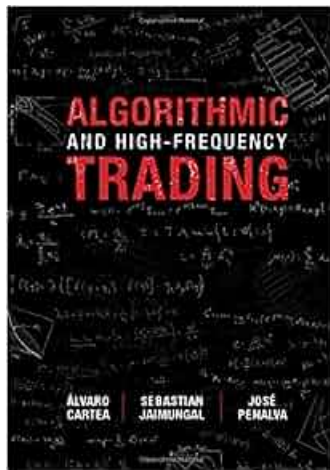
CoRR abs/2306.12806 (2023)

- ▶ While the CGAN we analysed had impressive **realism in isolation**
- ▶ We were able to construct hand-crafted and RL-based trading agents that **exploited its features and mechanism** to earn unrealistic profits
- ▶ That is, more work is needed on **interactive realism**
- ▶ This is an **important and exciting direction**

# Mathematical models

**Algorithmic and High-Frequency Trading (2015)**

**Álvaro Cartea, Sebastian Jaimungal, José Penalva**



Financial applications of Reinforcement Learning (RL)

Introduction to Limit Order Books

Types of Limit Order Book RL environment

- Market replay

- Conditional generative models

- Mathematical models

**Mathematical models of limit order book trading**

Tutorial on mbt-gym (model-based-trading gym)

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$



# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $\mathbf{S} = (S_t)_{t \in [0, T]}$

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $\mathbf{S} = (\mathbf{S}_t)_{t \in [0, T]}$  (AS: Brownian motion)

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $\mathbf{S} = (\mathbf{S}_t)_{t \in [0, T]}$  (AS: Brownian motion)
- ▶ **Arrival processes**  $\mathbf{M}^\pm = (\mathbf{M}_t^\pm)_{t \geq 0}$

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $S = (S_t)_{t \in [0, T]}$  (AS: Brownian motion)
- ▶ **Arrival processes**  $M^\pm = (M_t^\pm)_{t \geq 0}$  (AS: Poisson processes)

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $S = (S_t)_{t \in [0, T]}$  (AS: Brownian motion)
- ▶ **Arrival processes**  $M^\pm = (M_t^\pm)_{t \geq 0}$  (AS: Poisson processes)
- ▶ Conditional upon arrivals, limit orders are filled according to a **fill probability model**

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $S = (S_t)_{t \in [0, T]}$  (AS: Brownian motion)
- ▶ **Arrival processes**  $M^\pm = (M_t^\pm)_{t \geq 0}$  (AS: Poisson processes)
- ▶ Conditional upon arrivals, limit orders are filled according to a **fill probability model** (AS:  $\mathbb{P}[\text{fill} | \delta^\pm] = e^{-\kappa \delta^\pm}$ )

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $\mathbf{S} = (S_t)_{t \in [0, T]}$  (AS: Brownian motion)
- ▶ **Arrival processes**  $\mathbf{M}^\pm = (M_t^\pm)_{t \geq 0}$  (AS: Poisson processes)
- ▶ Conditional upon arrivals, limit orders are filled according to a **fill probability model** (AS:  $\mathbb{P}[\text{fill} | \delta^\pm] = e^{-\kappa \delta^\pm}$ )
- ▶ **Reward functions**

# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $S = (S_t)_{t \in [0, T]}$  (AS: Brownian motion)
- ▶ **Arrival processes**  $M^\pm = (M_t^\pm)_{t \geq 0}$  (AS: Poisson processes)
- ▶ Conditional upon arrivals, limit orders are filled according to a **fill probability model** (AS:  $\mathbb{P}[\text{fill} | \delta^\pm] = e^{-\kappa \delta^\pm}$ )
- ▶ **Reward functions** (AS:  $R_T = \exp(-\gamma \text{PnL}_T)$ )



# A family of LOB trading models

The AS model – the first of many related models

**High-frequency trading in a limit order book**

**Marco Avellaneda and Sasha Stoikov**

**Quantitative Finance 2008**

Recurring model structure

- ▶ Controls: bid and ask quote **depths**  $\delta^\pm = (\delta_t^\pm)_{t \in [0, T]}$
- ▶ **Midprice process**  $\mathbf{S} = (S_t)_{t \in [0, T]}$  (AS: Brownian motion)
- ▶ **Arrival processes**  $\mathbf{M}^\pm = (M_t^\pm)_{t \geq 0}$  (AS: Poisson processes)
- ▶ Conditional upon arrivals, limit orders are filled according to a **fill probability model** (AS:  $\mathbb{P}[\text{fill} | \delta^\pm] = e^{-\kappa \delta^\pm}$ )
- ▶ **Reward functions** (AS:  $\mathbf{R}_T = \exp(-\gamma \text{PnL}_T)$ )

# Some models covered by `mbt_gym`

Examples of models from the literature:

- ▶ Avellaneda and Stoikov [AS08]
- ▶ Market-making with limit orders [CJP15, Section 10.2]
- ▶ Market-making at the touch [CJP15, Section 10.2.2]
- ▶ Cartea, Jaimungal, and Ricci (CJR) [CJR14]
- ▶ Guéant, Lehalle, and Fernandez-Tapia [GLFT13] – section 5.2

Examples of extensions:

- ▶ Optimal execution with limit and market orders [CJP15, Section 8.4]
- ▶ Avellaneda and Stoikov with Hawkes' order flows
- ▶ [CJR14] with pure jump price process
- ▶ [CJR14] with exponential utility

... and many more!

# Solving these models

- ▶ For many of these models **optimal solutions** are known
- ▶ Found typically via the **Hamilton-Jacobi-Bellman equations** (whose solution is the value function of the control problem)
- ▶ RL can go wrong in many ways
- ▶ Having these optimal solutions to hand allows us to **compare the performance of RL in terms of both the reward (value) and the policy itself**

# Learning to make a market using PPO

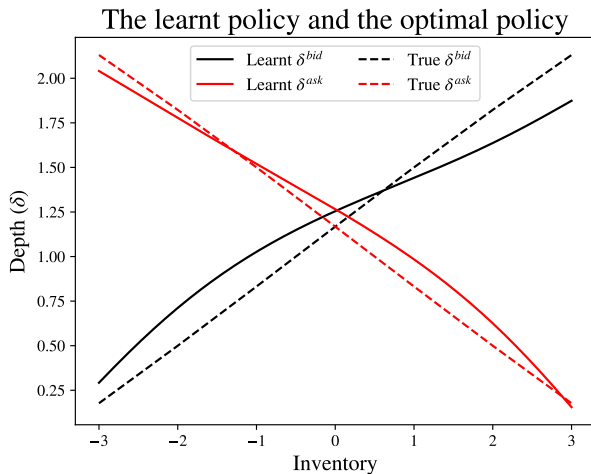






Figure: Learnt PPO agent versus CarteaJaimungalAgent.

## Ultimate vision

- ▶ Often model assumptions are driven by the **mathematical tractability** of finding optimal policies
- ▶ RL can be explored in contexts where optimal solutions are known (to make sure learning works well)
- ▶ Then **apply it settings where mathematical derivation is intractable**
- ▶ However, there are serious challenges on how one might use RL to effectively find (approximately) optimal solutions as a function of model parameters...

# References

-  Marco Avellaneda and Sasha Stoikov, *High-frequency trading in a limit order book*, Quantitative Finance **8** (2008), no. 3, 217–224.
-  Álvaro Cartea, Sebastian Jaimungal, and José Penalva, *Algorithmic and High-Frequency Trading*, Cambridge University Press, 2015.
-  Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci, *Buy low, sell high: A high frequency trading perspective*, SIAM Journal on Financial Mathematics **5** (2014), no. 1, 415–444.
-  Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez-Tapia, *Dealing with the Inventory Risk: A solution to the market making problem*, Mathematics and Financial Economics **7** (2013), no. 4, 477–507.

# mbt\_gym repo

The screenshot shows the GitHub repository page for JJJerome/mbt\_gym. At the top, there's a navigation bar with links to code, issues (1), pull requests, actions, projects, wiki, security (1), and insights. Below this, the repository name 'mbt\_gym' is displayed as 'Public'. To the right, there are buttons for 'Unwatch' (4), 'Fork' (18), and 'Starred' (88). Below the repository name, there are buttons for 'main' (selected), '3 branches', and '0 tags'. To the right of these are buttons for 'Go to file', 'Add file', and 'Code'. The main content area shows a list of files and folders with their commit history:

File/Folder	Commit Message	Commit Hash	Time
leandro-sbetancourt Merge pull request #35 from JJJerome/remove_max_inv_cj...		d9e75e1	last week
432 commits			
docker	Start time callable		6 months ago
experiments	Remove max inventory from CJMM agent		last week
mbt_gym	Remove max inventory from CJMM agent		last week
notebooks	back to 10,000,000 training		2 months ago
.gitignore	minor tidy up for Learning_Hawkes_arrivals.ipynb		3 months ago

On the right side, there is an 'About' section with the following text:

mbt\_gym is a module which provides a suite of gym environments for training reinforcement learning (RL) agents to solve model-based high-frequency trading problems such as market-making and optimal execution. The module is set up in an extensible way to allow the combination of different aspects of different models. It supports highly efficient ...

[https://github.com/JJJerome/mbt\\_gym](https://github.com/JJJerome/mbt_gym)

Financial applications of Reinforcement Learning (RL)

Introduction to Limit Order Books

Types of Limit Order Book RL environment

- Market replay

- Conditional generative models

- Mathematical models

Mathematical models of limit order book trading

Tutorial on mbt-gym (model-based-trading gym)



**Before we explore `mbt_gym`,  
any questions?**

[https://github.com/JJJerome/mbt\\_gym](https://github.com/JJJerome/mbt_gym)

[https://github.com/JJJerome/mbt\\_gym](https://github.com/JJJerome/mbt_gym)

For the **two notebooks** we will now go through go to branch:  
**oxml**

