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

Outline

Applications of Reinforcement Learning in Finance

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

Excellent survey

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)

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**

Bids	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	

Best prices, midprice, and spread

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

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

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

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?

	Bids	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		

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

How do they make money?

requires others to **"cross the spread"** in **both** directions, i.e.,

someone lifts MM bid; someone **else** lifts MM ask

Market order example

Bids	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	

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

Bids	Price	Asks	
	101.50	5	
	101.25	5	
	101.00		last MM sale
	100.50		
	100.25		
	100.00		
14	99.75		old MM bid
30	99.50		
3	99.25		
11	99.00		

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

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

Midprice goes up

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

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

Toxic order flow and inventory risk

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

Inventory risk: may amass undesired position in trending markets

Toxic order flow and inventory risk

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

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

Bids	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	

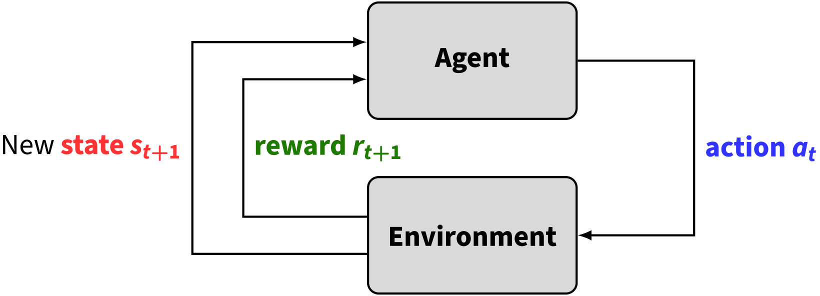
Reinforcement Learning from scratch through live trading is generally infeasible; some form of **simulated environment** is used

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)

Reinforcement learning environment



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

LOB environments: Three main possiblities

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

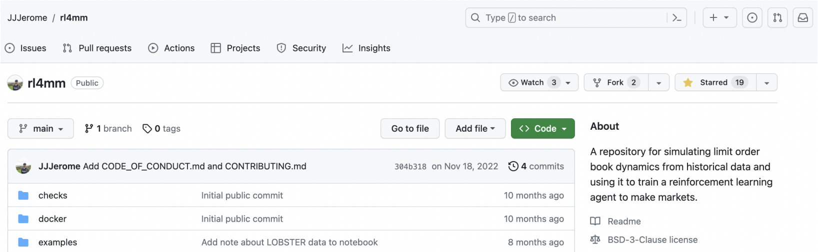
but, evidence shows they tend to happen near back of queue

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)



`https://github.com/JJJerome/rl4mm`

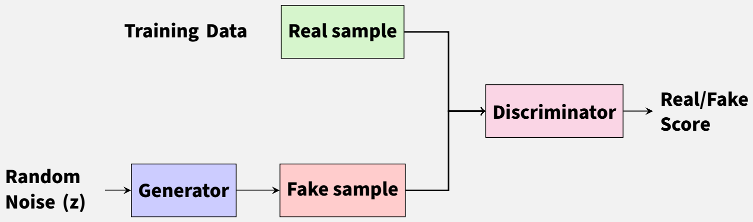
Conditional Generative Model as the simulator

- 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**

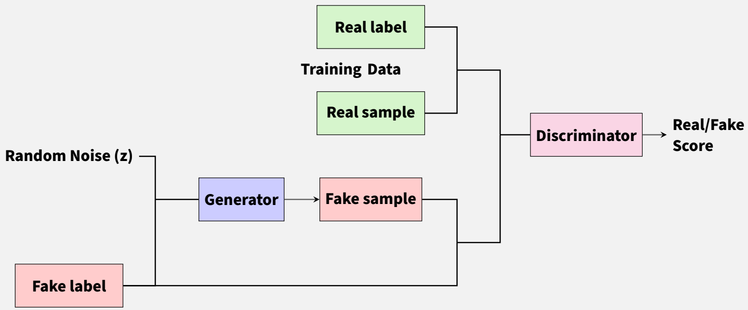
(unconditional) Generative Adversarial Network (GAN)



Conditional Generative Model as the simulator

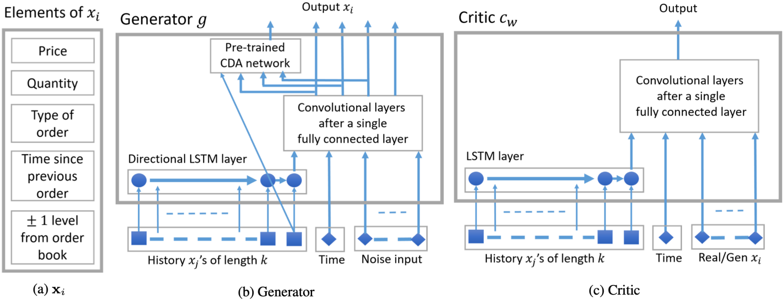
- Simulator based on **conditional generative model** that was trained on real order book data
- **Promises both realism and reactivity**

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 Vyetrenko

CoRR abs/2306.12806 (2023)

More work to be done

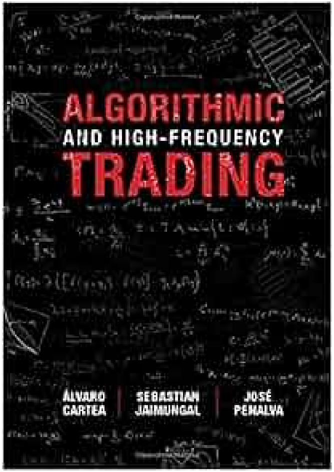
Conditional Generators for Limit Order Book Environments: Explainability, Challenges, and Robustness

Andrea Coletta, Joseph Jerome, Rahul Savani, Svitlana Vyetrenko

CoRR abs/2306.12806 (2023)

- While the CGAN we analysed had impressive **realism in isolation**
- We we 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**

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** $S = (S_t)_{t\in[0,T]}$

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** $S = (S_t)_{t \in [0,T]}$ (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** $S = (S_t)_{t \in [0,T]}$ (AS: Brownian motion)
- **Arrival processes** $M^{\pm} = (M_t^{\pm})_{t \geq 0}$

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** $S = (S_t)_{t \in [0,T]}$ (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**

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: $R_T = \exp\left(-\gamma \mathrm{PnL}_T\right)$)

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: $R_T = \exp\left(-\gamma \mathrm{PnL}_T\right)$)

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

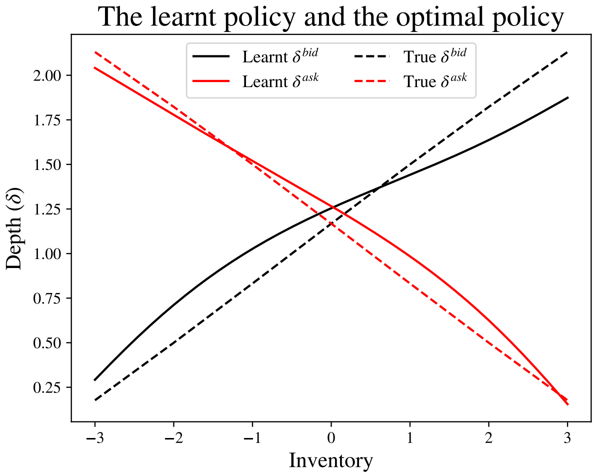


Figure: Learnt PPO agent versus `CarteaJaimungalAgent`.

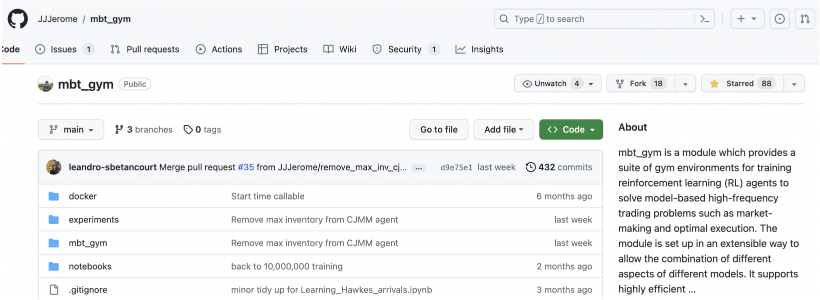
Ultimate vision

- 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



https://github.com/JJerome/mbt_gym

Before we explore `mbt_gym`, any questions?

mbt_gym repo

https://github.com/JJJerome/mbt_gym

mbt_gym repo

https://github.com/JJJerome/mbt_gym

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