Limit Order Book Simulation:
A Review and Recent Progress

Ahmer Nadeem Khan

Financial Math Seminar 2025

Decmber 4, 2025

Table of Contents

Table of Contents

Trading Venues

Definition (Financial Markets)

Financial Markets are venues for buying and selling securities.

- Traditional exchanges have centralized limit order books (CLOBs).
- Equities (NYSE, NASDAQ), options (Cboe), futures (CME), and most centralized crypto markets.
- Differ in tick sizes, lot sizes, queuing rules, and participant composition.

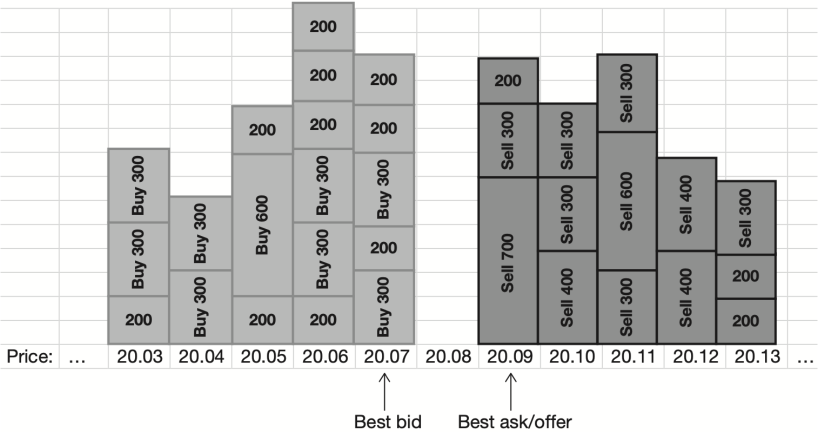
Limit Orders

Definition (Limit Order)

A *limit order* is an order to buy or sell a stock with a restriction on the maximum price to be paid or the minimum price to be received.

- The **best bid** is the highest buy limit available.
- The **best ask/offer** is the lowest market sell limit available.
- Passive order; sits in the LOB. Larger and more limit orders means more liquidity.

Limit Orders



Market Orders

Definition (Market Order)

A *market order* is an order to buy or sell a stock at the market's best available price.

- Market **buy orders** meet best ask/offer.
- Market **sell orders** meet best bid.
- Active orders placed during market hours.

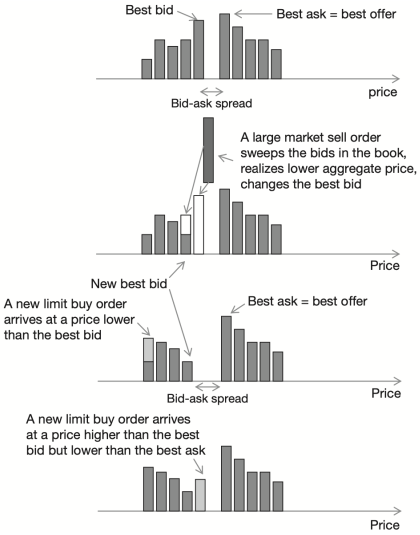
Limit Order Book

Definition (Limit Order Book)

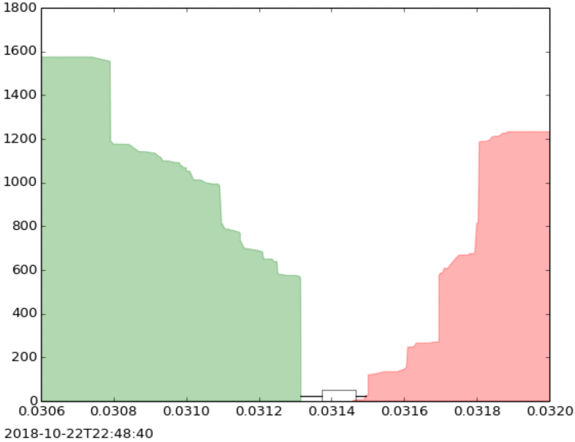
A *limit order book* is a real-time record of all limit buy and sell orders in a market, organized by price levels.

- Any new limit order is placed into a limit order queue corresponding to its price.
- The queue is generally *price-time* priority (FIFO).
- Another important order type is 'Cancel' - limit orders may be canceled.

Limit Order Book Dynamics [1]



Bittrex ETH/BTC order book [2]



- x-axis is the unit price of Ethereum in terms of Bitcoin.
- y-axis is cumulative order depth in Bitcoin.

Trader Screen [3]

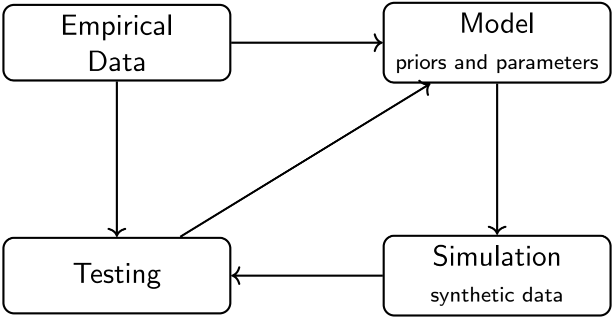
100		11	22.450	**22.71**	23.000	11	**26**		50	0.013	**0.04**	0.075	198		0.002	0.000
99		11	20.475	**20.73**	21.000	11	**28**	-1	153	0.050	**0.08**	0.100	115		0.007	0.000
97		11	18.525	**18.79**	19.075	12	**30**	-163	152	0.125	**0.15**	0.175	134		0.015	0.000
95		12	16.625	**16.90**	17.175	12	**32**	-64	149	0.225	**0.26**	0.275	1		0.025	0.000
94		12	15.700	**15.97**	16.250	12	**33**		92	0.300	**0.34**	0.375	149		0.030	0.000
93		12	14.775	**15.06**	15.325	12	**34**	29	27	0.400	**0.43**	0.475	152	3	0.036	0.000
91		12	13.900	**14.17**	14.425	12	**35**	-99	43	0.500	**0.53**	0.575	133	20	0.042	0.000
90		12	13.025	**13.30**	13.550	13	**36**	-118	230	0.600	**0.66**	0.700	232	2	0.048	0.000
88		13	12.150	**12.45**	12.700	13	**37**	38	178	0.750	**0.81**	0.850	208		0.054	0.000
86		13	11.325	**11.62**	11.875	13	**38**	-11	122	0.925	**0.97**	1.000	124	2	0.061	0.000
84		13	10.525	**10.82**	11.050	14	**39**	150	146	1.100	**1.16**	1.225	180	16	0.068	-0.00
82		13	9.750	**10.04**	10.275	14	**40**	-50	33	1.325	**1.38**	1.400	47	22	0.074	0.000
79		25	9.075	**9.28**	9.500	27	**41**	101	114	1.550	**1.62**	1.650	95	2	0.080	0.000
77		26	8.325	**8.55**	8.775	28	**42**	94	27	1.825	**1.88**	1.925	68	6	0.086	0.000
74		26	7.650	**7.85**	8.050	15	**43**	89	35	2.125	**2.18**	2.225	77	12	0.092	0.000
72	3	27	6.975	**7.18**	7.375	16	**44**	34	29	2.450	**2.50**	2.550	63	217	0.097	0.000
69		29	6.325	**6.53**	6.725	17	**45**	175	19	2.775	**2.85**	2.900	61	4	0.102	-0.00
66	3	15	5.750	**5.92**	6.075	17	**46**	15	27	3.150	**3.23**	3.275	54	20	0.105	0.000
63		27	5.225	**5.33**	5.425	30	**47**	57	61	3.550	**3.64**	3.675	51		0.108	0.000
60	26	39	4.675	**4.78**	4.850	34	**48**	-55	16	4.000	**4.08**	4.150	64	240	0.111	0.000
56	29	41	4.150	**4.26**	4.325	23	**49**	-149	36	4.450	**4.55**	4.600	45		0.112	0.001
53	36	37	3.675	**3.78**	3.825	26	**50**	-88	39	4.950	**5.06**	5.150	55	2	0.112	0.001
50	11	40	3.225	**3.32**	3.400	59	**51**	-25	27	5.500	**5.60**	5.650	36	1	0.112	0.000
46	8	43	2.800	**2.90**	2.950	32	**52**	-276	32	6.075	**6.17**	6.275	54	17	0.110	0.002
42	8	59	2.425	**2.52**	2.550	38	**53**	-98	18	6.600	**6.78**	6.925	18		0.108	0.001
39	6	10	2.100	**2.16**	2.200	36	**54**	-145	18	7.250	**7.42**	7.575	18		0.104	0.001
35	3	85	1.775	**1.85**	1.900	63	**55**	-116	17	7.925	**8.10**	8.300	33	2	0.100	0.001
31	101	98	1.500	**1.56**	1.625	64	**56**	-129	16	8.675	**8.81**	8.975	16	22	0.095	0.001
27	1	110	1.250	**1.31**	1.375	94	**57**	35	15	9.375	**9.56**	9.775	29		0.089	0.000
24	6	101	1.050	**1.10**	1.150	62	**58**	81	15	10.125	**10.34**	10.550	27		0.083	-0.00
20		113	0.875	**0.92**	0.975	108	**59**	3	28	10.925	**11.15**	11.375	27		0.076	0.000
17		20	0.775	**0.77**	0.825	141	**60**	238	14	11.775	**11.99**	12.275	14		0.070	-0.00
14		185	0.600	**0.64**	0.675	106	**61**	-4	14	12.650	**12.87**	13.150	14		0.064	0.000
12	124	187	0.500	0.53	0.575	132	62	123	14	13.525	13.76	14.050	13		0.058	-0.00
8		187	0.325	**0.37**	0.425	159	**64**	-141	13	15.350	**15.59**	15.900	13		0.047	0.001
6	5	151	0.225	**0.27**	0.300	115	**66**	-17	13	17.225	**17.48**	17.775	12		0.037	0.000

Figure: Options market-maker trading screen.

Table of Contents

Modeling Challenges

- LOB data is very difficult to model and simulate accurately and efficiently.
- Decades of empirical study has enabled researchers to identify statistical properties of LOBs.



Applications

- Real market data has $n = 1$ (statistical issues).
- Synthetic data is used to train reinforcement learning (RL) agents for many purposes.
- Market making, statistical arbitrage, optimal execution, out-of-sample testing.
- Conduct stress-tests for portfolio or trading strategies.

Market Impact

- We don't want a simple replay of the history.
- Response to exogenous trades by autonomous agents.
- This reaction ability is called *market impact awareness*.
- **Example**: learning optimal execution policy.

Stylized Facts

Let b_t, a_t denote the best bid and best ask prices respectively. The mid-price is defined as $m_t = \frac{b_t - a_t}{2}$, and the (log) return is defined by $r_{t,\Delta t} = \log(m_{t+\Delta t}) - \log(m_t)$.

The following are stylized facts about *asset return distributions*.

- **Linear autocorrelations** $\mathrm{corr}(r_{t+\tau,\Delta t}, r_{t,\Delta t})$ for a period τ are insignificant for $\tau > 15$ minutes.
- **Volatility clustering** as measured by $\mathrm{corr}(r^2_{t+\tau,\Delta t}, r^2_{t,\Delta t})$. This remains positive over several days.
- **Long-range dependence** decays by a power law: $f(\tau) = \mathrm{corr}(|r_{t+\tau,\Delta t}|, |r_{t,\Delta t}|) \sim \tau^{-\beta}$ with exponent $\beta \in [0.2, 0.4]$.

Stylized Facts (Cont.)

- **High volatility** over micro or macro time scales.
- **Gain/Loss asymmetry** as measured by skewness. Stocks lose value faster than they grow.
- **Heavy tails** as measured by kurtosis, which become slimmer with longer periods.
- **Volume/Volatility** have positive correlation.
- **Returns/Volatility** have negative correlation.

Stylized Facts (Cont.)

We can also find similar results for *volumes and order flows*.

- **Quantities of interest** are quote volumes, quote sizes, number of quotes in a fixed time window, inter-arrival times, new quote prices, order lifetimes (e.g. cancellation times) etc.
- **Parametric distributions** used to fit are usually power-law, exponential, gamma, lognormal, and Weibull distributions etc.

Finally, we also consider *non-stationary patterns*.

- Intraday volume/spread negative correlation.
- Intraday sensitivity to macro economic events/holidays.
- Intraday and seasonal volume patterns.

Important anomalies: *cross-asset correlation*.

Research

Earliest work is from 1999-2002 e.g.

- Bouchaud, J.-P., Mézard, M., & Potters, M. (2002). *Statistical properties of stock order books: empirical results and models*. Quantitative Finance [4]
- Cont, R. (2001). *Empirical properties of asset returns: stylized facts and statistical issues*. Quantitative Finance [5].

Currently an active area of research. A recent survey is

- Vyetrenko, S., Byrd, D., Petosa, N., Mahfouz, M., Dervovic, D., Veloso, M., & Balch, T. (2021). *Get real: Realism metrics for robust limit order book market simulations*. In Proceedings of the First ACM International Conference on AI in Finance (ICAIF '20) [6].

Table of Contents

Modeling the LOB

We distinguish three categories of orders:

- **Limit orders**: buy (resp. sell) orders placed at prices less than or equal to the best ask (resp. greater than or equal to the best bid).
- **Market orders**: orders that immediately execute against the best available prices in the order book.
- **Cancel orders**: cancellations of outstanding limit orders without execution.

Order Book State

Definition (Limit Order Book State)

The state of the limit order book at time t is defined as

$$\boldsymbol{S}_t := \big(B_t, A_t, v_{b,t}(x), v_{a,t}(x)\big),$$

where:

- B_t and A_t denote the best bid and best ask prices,
- $v_{b,t}(x)$ (resp. $v_{a,t}(x)$) denotes the bid-side (resp. ask-side) volume at price level x,
- x represents the distance in price units from the mid-price P_t, with

  $$x \in n^*\mathbb{Z}, \qquad v_{b,t}(x),\ v_{a,t}(x) \geq 0,$$

  where n^* denotes the minimum tick size.

Order Book Dynamics

The evolution of the order book is driven by 8 types of events:

- Market orders (buy/sell),
- In-spread limit orders (bid/ask),
- Cancel orders (bid/ask),
- Limit orders outside the spread (bid/ask).

The event-level dynamics are given by

$$\boldsymbol{S}_{t+1} = \boldsymbol{S}_t + \mathcal{D}_t(\boldsymbol{S}_t),$$

where $\mathcal{D}_t(\cdot)$ is a random operator which depends on the dynamics that each 8 kinds of events. The Order Book state is partially observable and is often modeled as a Markovian system.

Point Processes

Assume that the LOB is a queue system with different orders arriving independently.

Let $N_{t,t+\Delta t}$ denote the number of events occurring in the time interval $(t, t+\Delta t]$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $\{\mathcal{F}_t\}_{t\geq 0}$ the natural filtration.

The conditional intensity of the point process is defined as

$$\lambda_t := \lim_{\Delta t \to 0} \frac{\mathbb{P}(N_{t,t+\Delta t} > 0 \mid \mathcal{F}_t)}{\Delta t},$$

$$\mathbb{P}\big(N_{t,t+\Delta t} > 0 \mid \mathcal{F}_t\big) = \lambda_t \, \Delta t + o(\Delta t).$$

Poisson Point Process

Definition (Poisson Process)

A point process $\{N_{t,t+\Delta t}\}$ is called a *Poisson process* with rate $\lambda > 0$ if it satisfies the following properties:

1. For any interval $(t, t+\Delta t]$,

$$N_{t,t+\Delta t} \sim \mathrm{Poisson}(\lambda \cdot \Delta t).$$

2. For any two non-overlapping intervals $(t, t+\Delta t]$ and $(s, s+\Delta s]$, the random variables

$$N_{t,t+\Delta t} \quad \text{and} \quad N_{s,s+\Delta s}$$

are independent.

Zero-Intelligence (ZI) Models

- The **zero-intelligence models** describe the market without informed, heterogeneous decision making agents, e.g. the Poisson process with constant intensity.
- Reveal the order matching dynamics of the LOB.
- Relax conditions; stationarity, multi-dimensional, dependence on variables e.g. queue size, or distance from mid-price etc.
- Analytically tractable, and well-developed theory.

A Poisson Process Model

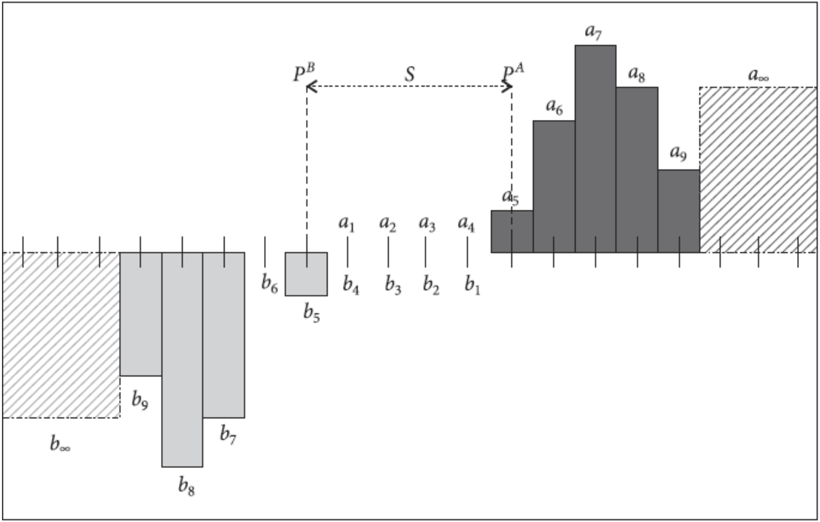


Figure: Model in *Limit Order Books, Abergel (2016)* [7]

Poisson Process Example

- For the ask and bid sides of the book, define the cumulative volumes

$$A_i := \sum_{k=1}^{i} a_k, \qquad B_i := \sum_{k=1}^{i} |b_k|,$$

- For a given quantity of shares q', define

$$A^{-1}(q') := \inf\{\, p : A_p > q' \,\}, \qquad B^{-1}(q') := \inf\{\, p : B_p > q' \,\}.$$

- The index of the first non-empty limit on both sides of the book is denoted by i_S and is given by

$$i_S := A^{-1}(0) = B^{-1}(0) = \frac{S}{\Delta P},$$

where S is the bid–ask spread and ΔP is the tick size.

- Boundary conditions (liquidity reservoirs) ensure that $i_S < \infty$.

Order Arrivals and Notation

The arrival of events are modeled by independent Poisson processes:

- $M^{\pm}(t)$: market orders (buy/sell) with constant intensities $\lambda^{M^{\pm}}$;
- $L_i^{\pm}(t)$: limit orders at level i with constant intensities $\lambda_i^{L^{\pm}}$;
- $C_i^{\pm}(t)$: cancellations at level i, with stochastic intensities $\lambda_i^{C^{\pm}}|a_i|$, proportional to queue size.

where

- Superscript "$+$" (resp. "$-$") denotes ask (resp. bid) side;
- All orders have fixed size q;
- Cancellation intensity is zero whenever the corresponding queue is empty.

Stochastic Differential Equations (SDEs)

Let $a_i(t)$ denote the number of shares available i ticks above the best bid on the ask side. The evolution of $a_i(t)$ is given by the stochastic differential equation

$$\begin{aligned} da_i(t) = -\mathbf{1}_{\{a_i(t)\neq 0\}}(q - A_{i-1})_+ \, dM^+(t) + q \, dL_i^+(t) - q \, dC_i^+(t) \\ + (J^{M^-}(a) - a)_i \, dM^-(t) + \sum_{j=1}^{K} (J_j^{L^-}(a) - a)_i \, dL_j^-(t) \\ + \sum_{j=1}^{K} (J_j^{C^-}(a) - a)_i \, dC_j^-(t). \end{aligned}$$

We can write a similar equation for $db_i(t)$

SDEs (Cont.)

- We can then write down the *infinitesimal generator* associated with this $2K$-dimensional continuous-time Markov chain.
- Then, determine the SDEs for the price and spread dynamics.
- It can be shown that that, under strong intensity of cancellations, the Limit Order Book is an **ergodic** Markov process, with exponential rate of convergence to a unique stationary distribution.
- Furthermore, there is a large scale limit of the process which can be expressed in terms of **Brownian motion**.
- Also an active area of research [8].

Hawkes Process

- We can relax the assumption of independence and introduce cross and self excitation terms.
- for e.g., a 2-D Hawkes Process of (ask volume, bid volume) can have 4 excitation terms:
  $ask \rightarrow ask, bid \rightarrow bid, ask \rightarrow bid, bid \rightarrow ask$.
- In general, for a d-dimensional Hawkes process, the intensity of the process $\lambda_t^{(i)}$ and the associated counting process $N_t^{(i)}$ for $i = 1, \dots, d$ is defined as

$$\lambda_t^{(i)} = \mu_t^{(i)} + \sum_{j=1}^{d} \int_0^t \phi^{(j \rightarrow i)}(t-s)\, dN_s^{(j)}.$$

Agent-based Models (ABMs)

- Markets consist of autonomous agents; market heterogeneity.
- Cont, Cucuringu, Glukhov, et al. (2023) [9] used clustering techniques to show that there are at least 4 different clusters of traders.
- High frequency traders, trend followers, mean reverters, noise traders and algorithmic traders.
- Mixed models and game-theoretic models.

ABM Example 1 (Herding) [7]

- Assume N traders, with demand $\phi_i(t) = \pm 1$ and probability $a \in (0,1)$ of acting.
- The price is assumed to have the form:

$$p(t+1) = p(t) + \frac{1}{\lambda}\sum_{i=1}^{N}\phi_i(t)$$

where λ is the market depth.

- We can begin clustering and linking traders:

$$\Delta p(t) = \frac{1}{\lambda}\sum_{i=1}^{n_c(t)} W_k \phi_k$$

where $n_c(t)$ is the number of clusters and W_k is the size of cluster k.

ABM Example 2

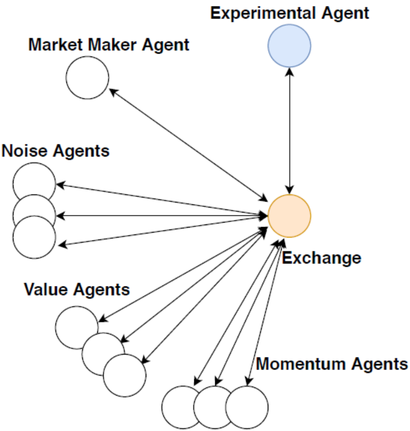


Figure: Agent-based model in *Vyetrenko (2020)* [6]

Deep Learning Models

- Complexities and non-nonlinearities in the time evolution and distributions.
- Use neural networks to predict mid-price, the volatility and the direction etc [8].
- Popular architectures being used are Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs) [10].

Large Language Models

- Nagy, B. et al. (2023). *Generative AI for End-to-End Limit Order Book Modelling: A Token-Level Autoregressive Generative Model of Message Flow Using a Deep State Space Network. In Proceedings of the Fourth ACM International Conference on AI in Finance (ICAIF '23) [11].*
- Treat order flow like natural language.
- Generate entire message streams.
- Reconstruct full LOB trajectories via a simulator [12].

LLMs (Cont.)

- Used NASDAQ LOBSTER data.
- Architecture: deep state space model composed of simplified structured state space layers (S5).

$$
\begin{aligned}
\dot{x}(t) &= A\,x(t) + B\,u(t), \\
y(t) &= C\,x(t) + D\,u(t),
\end{aligned}
$$

- A **state-space model** represents a sequence by evolving a latent memory state that continuously integrates past inputs and produces predictions through a learned dynamical system.
- Autoregressive: (conditional) probability distribution over entire sequences of tokens.
- Use perplexity (PPL) to evaluate the probability the model assigns to the test data

LLMs (Cont.)

- P separate volume features around the mid-price, and one feature for changes in mid-price.
- Book data (state) is a feature.
- Tokens for: Event-type, Price (sign, distance), Size, Inter arrival time, Referential $\approx$ 12000 total vocabulary size.
- The problem becomes

$$\min_{\theta} \frac{1}{N} \sum_{\{m,b,y\}} -\log \hat{y}_{\{i=y\}}, \quad \text{with } \hat{y} = f_{\theta}(m, b).$$

  where m, b are the message and book sequences
- Match statistical properties well, low perplexity score, and high-correlation with mid-price series.

Tokenization

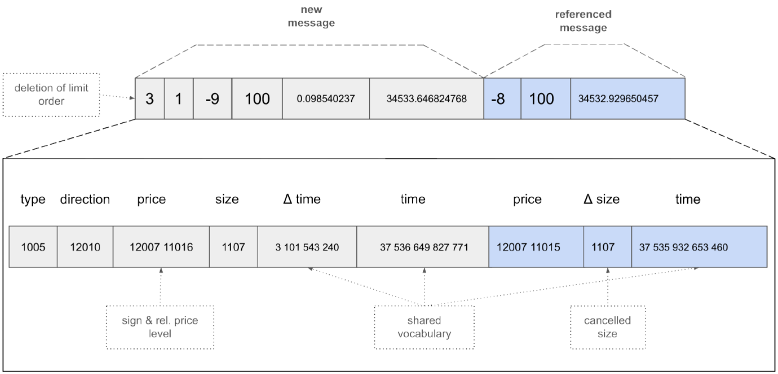


Figure: Encoding example in Nagy et al.

Table of Contents

Conclusions

- Limit Order Book simulation is an active area of research, both theoretically and practical.
- High-fidelity and efficient simulation has vast applications in finance, in particular trading.

Table of Contents

References I

Irene Aldridge.
High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems.
John Wiley & Sons, Hoboken, NJ, 2 edition, 2013.
Wiley Trading Series.

Kjerish.
Bittrex eth/btc order book depth chart.
https://commons.wikimedia.org/wiki/File:Order_book_depth_chart.gif, October 2018.
Licensed under CC BY-SA 4.0.

References II

Akuna Capital.
Sample trading order book screen for options market making.
Online course material, n.d.
From Akuna Capital, Options 101 course. Image obtained via public course content and used for illustrative purposes.

Jean-Philippe Bouchaud, Marc Mézard, and Marc Potters.
Statistical properties of stock order books: Empirical results and models.
December 2002.
Available at SSRN.

R. Cont.
Empirical properties of asset returns: stylized facts and statistical issues.
Quantitative Finance, 1(2):223–236, 2001.

References III

Svitlana Vyetrenko, David Byrd, Nick Petosa, Mahmoud Mahfouz, Danial Dervovic, Manuela Veloso, and Tucker Balch.

Get real: realism metrics for robust limit order book market simulations.
In *Proceedings of the First ACM International Conference on AI in Finance*, ICAIF '20, New York, NY, USA, 2021. Association for Computing Machinery.

Frédéric Abergel, Marouane Anane, Anirban Chakraborti, Aymen Jedidi, and Ioane Muni Toke.
Limit Order Books.
Physics of Society: Econophysics and Sociophysics. Cambridge University Press, Cambridge, 1st edition, 2016.

References IV

Konark Jain, Nick Firoozye, Jonathan Kochems, and Philip Treleaven.
Limit order book simulations: A review.
Papers 2402.17359, arXiv.org, Feb 2024.

Rama Cont, Mihai Cucuringu, Vacslav Glukhov, and Felix Prenzel.
Analysis and modeling of client order flow in limit order markets.
Quantitative Finance, 23(2):187–205, 2023.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
Deep Learning.
MIT Press, 2016.
http://www.deeplearningbook.org.

References V

Peer Nagy, Sascha Frey, Silvia Sapora, Kang Li, Anisoara Calinescu, Stefan Zohren, and Jakob Foerster.
Generative ai for end-to-end limit order book modelling: A token-level autoregressive generative model of message flow using a deep state space network.
In *Proceedings of the Fourth ACM International Conference on AI in Finance*, ICAIF '23, page 91–99, New York, NY, USA, 2023. Association for Computing Machinery.

Sascha Yves Frey, Kang Li, Peer Nagy, Silvia Sapora, Christopher Lu, Stefan Zohren, Jakob Foerster, and Anisoara Calinescu.
Jax-lob: A gpu-accelerated limit order book simulator to unlock large scale reinforcement learning for trading.

References VI

In *Proceedings of the Fourth ACM International Conference on AI in Finance*, ICAIF '23, page 583–591, New York, NY, USA, 2023. Association for Computing Machinery.