2024 JSM Data Driven Methods in Financial Markets

Attention-Based Reading, Highlighting, and Forecasting of the Limit Order Book

Jiwon Jung
Ph.D. Student
Department of Statistics
Purdue University

PURDUE UNIVERSITY

August 5, 2024

Table of Contents

Introduction

- Market Microstructure
- Limit Order Book Dynamics
- Previous Studies

Market Microstructure

Market Microstructure

Market Microstructure

Market Microstructure

- **macro vs. micro** $\leftarrow$ earth from the space vs. surface of the land

Market Microstructure

- **macro vs. micro** $\leftarrow$ earth from the space vs. surface of the land
  - The macro-movement refers to daily, weekly, and monthly movements of asset prices, and the micro-movement refers to the transactional (trade-by-trade) price behavior.
  - In market microstructure, we can no longer ignore high-frequency phenomena such as microstructure noises, liquidity, and transaction costs.

Market Order vs. Limit Order

Market Order (MO)

An order to buy/sell an asset at the best available price

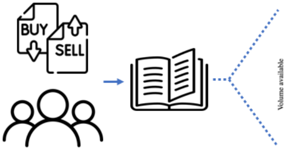
- pros: immediate execution, (i.e., no execution risk),
- cons: pay spread and an additional fee (f), liquidity risk.

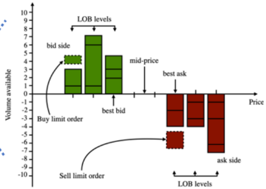
Limit Order (LO)

An order to buy/sell an asset at a specified price and quantity

- pros: get better price (when executed) and an additional rebate (r),
- cons: execution risk (the order may never be executed).

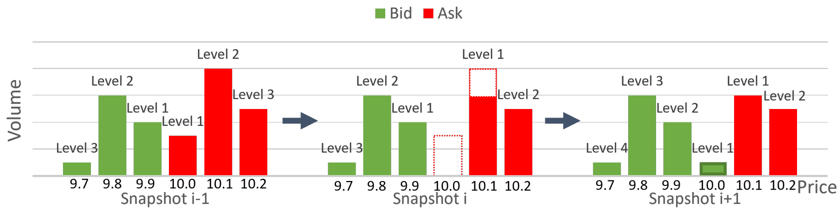
Limit Order Book (LOB)





- LOB: Price & time prioritized collection of buy and sell quotes.
- FIFO (first-in-first-out) policy for orders on the same price level.

Limit Order Book (LOB)



	price	size
Ask	10.2	500
	10.1	800
	10.0	300
Bid	9.9	400
	9.8	600
	9.7	100

→

	price	size
Ask	10.3	200
	10.2	500
	10.1	600
Bid	9.9	400
	9.8	600
	9.7	100

→

	price	size
Ask	10.3	200
	10.2	500
	10.1	600
Bid	10.0	100
	9.9	400
	9.8	600

Figure 1. A market buy order with size 500 and a limit bid order with size 100 at 10.0 in sequence.

Backgrounds: Transformer I

- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) were pioneers in sequence processing tasks.
  - Sequential Processing: RNN/LSTM process sequences step by step, limiting speed.
  - Vanishing Gradient Issue: RNNs often face vanishing gradient problems, hindering learning long-range dependencies.
- The Transformer architecture, introduced in "Attention Is All You Need" (Vaswani et al., 2017), reshaped sequence-to-sequence learning tasks, especially for NLP.
  - The attention mechanism in the Transformer looks at an input sequence and decides at each step which other parts of the sequence are important.
  - Parallelism: Transformer employs parallel processing, accelerating training time.

Backgrounds: Transformer II

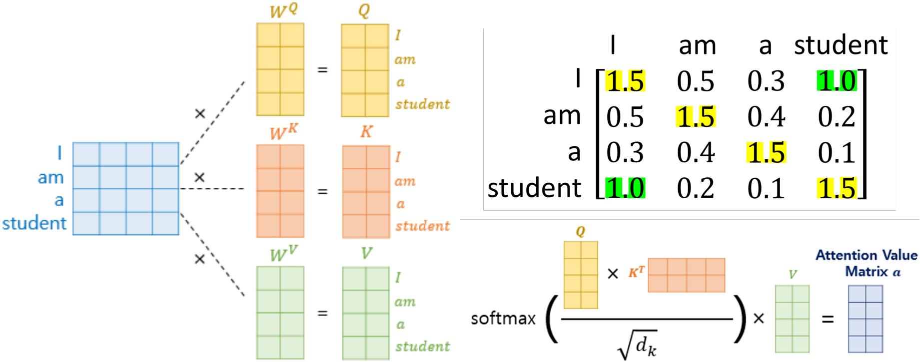


Figure 2. Illustration of Attention mechanism[1]

[1] credit: https://codingopera.tistory.com/43)

Previous Studies I

Mid-price forecasting:

- Ntakaris et al. (2018): regression-based approach for high-frequency limit order markets' mid-price prediction.
- Kercheval and Y. Zhang (2015): support vector machines for modeling whether a stock's mid-price increases, decreases, or remains the same;
- Sirignano (2019): a spatial neural network approach to capture the joint distribution of the state of the limit order book at a future time conditional on the current state of the limit order book.
- Z. Zhang et al. (2019): *DeepLOB* architecture utilizes convolutional filters to capture the spatial structure of the LOBs as well as long short-term memory modules to capture longer time dependencies.

Previous Studies II

Optimal execution:

- Park and Van Roy (2015) and Nevmyvaka et al. (2006): machine learning approaches (reinforcement learning) for optimal order execution

Motivation

- Limit order book has a unique and complicated structure: price, volume, level, time features
- Due to this complexity, many studies have simplified LOB prediction problems into mid-price forecasting/classifications.
- However, mid-prices have partial information on the LOB dynamics
  - BAS, depths, and levels that affect the dynamics are ignored.

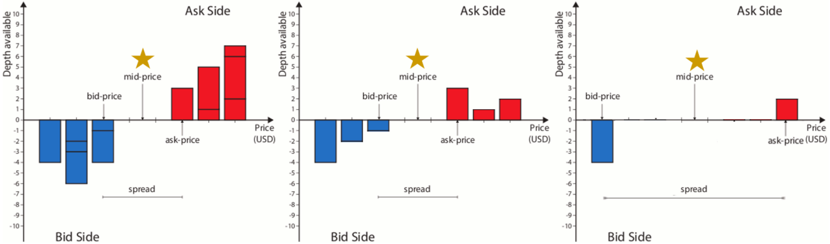


Figure 3. Three different LOB snapshots with the same mid-price (Figure modified from Fernandez-Tapia (2015))

Our Goal

- Our aim is to explore and predict the dynamic nature of limit order book data in its entirety, without simplifying the analysis solely to mid-price movements.

Attention-Based Multi-level Limit Order Book Analysis

LOBSTER Dataset

			Level 1				Level 2				...
			Ask		Bid		Ask		Bid		
Timestamp	Mid-price	Spread	Price	Quantity	Price	Quantity	Price	Quantity	Price	Quantity	
1275386347944	126200	200	126300	300	126100	17	126400	4765	126000	2800	...
1275386347981	126200	200	126300	300	126100	17	126400	4765	126000	2800	...
1275386347981	126200	200	126300	300	126100	17	126400	4765	126000	2800	...
1275386348070	126050	100	126100	291	126000	2800	126200	300	125900	1120	...
1275386348070	126050	100	126100	291	126000	2800	126200	300	125900	1120	...
1275386348101	126050	100	126100	291	126000	2800	126200	300	125900	1120	...

- LOBSTER[2]
  - A single day data from 5 different tech-stocks are publicly available
  - Timestamps are available.

[2] https://lobsterdata.com/

Unique Structure of Limit Order Book Data

- Spatial features: compound structure of multi-dimension
  - Type of orders (i): Ask & Bid
  - Features (j): Price & Volume[3]
  - Levels (k): $k = 1, \ldots, K$ (ex: $K = 1, 3, 5$, or 10)
  - Stocks (l): tickers (ex: AAPL, AMZN, GOOG, ...)
- Temporal features: order priority based on time of arrival

[3] Depth of market, displaying liquidity at different price levels

Problem Formulation I

We use $\vec{B}_i$ to denote the vector of the i^{th} bid orders and $\vec{A}_i$ to denote that of the i^{th} ask orders for each snapshot $i \in \{1, \ldots, N\}$ for each level $k \in \{1, \ldots, K\}$. Specifically,

$$\vec{B}_i = \left[t_i, (p^b_{1i}, v^b_{1i}), \ldots, (p^b_{Ki}, v^b_{Ki})\right], \tag{2.1}$$

where t_i is the time at which snapshot i was captured, p^b_{ki} is the level-k bid price, and $v^b_{ki}(t)$ is the level-k bid volume. Similarly, an ask order is represented as

$$\vec{A}_i = [t_i, (p^a_{1i}, v^a_{1i}), \ldots, (p^a_{Ki}, v^a_{Ki})], \tag{2.2}$$

where t_i is the time at which snapshot i was captured, p^a_{ki} is the level-k ask price, and $v^a_{ki}(t)$ is the level-k ask volume.

Spatiotemporal Embedding

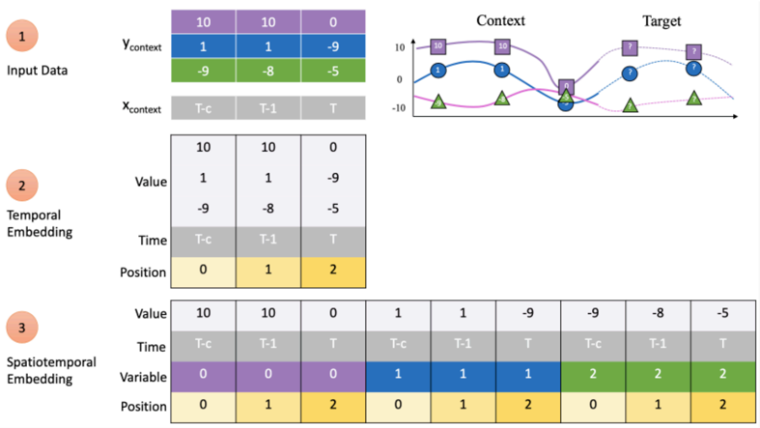


Figure 4. credit: Grigsby et al. (2021)

Attention-based LOB Reader

A Spatiotemporal sequence encoding (Spacetimeformer)

	AAPL level-1-ask-price	AAPL level-1-ask-volume	AAPL level-1-bid-price	AAPL level-1-bid-volume	AAPL level-2-ask-price	AAPL level-2-ask-volume	AAPL level-2-bid-price	AAPL level-2-bid-volume	...	Timestamp
Value	$585.68	900	$585.50	18	$224.45	100	$223.60	15	...	2012-06-21 09:30:05
Spatiotemporal embedding	0	1	2	3	4	5	6	7	...	

B Compound multivariate sequence encoding (ours)

		AAPL								...	Timestamp
		Level-1				Level-2				...	
		Ask		Bid		Ask		Bid		...	
		Price	Volume	Price	Volume	Price	Volume	Price	Volume	...	
Value		$585.68	900	$585.50	18	$224.45	100	$223.60	15	...	2012-06-21 09:30:05
Compound multivariate embedding	Type (i)	0	0	1	1	0	0	1	1	...	
	Feature (j)	0	1	0	1	0	1	0	1	...	
	Level (k)	0	0	0	0	1	1	1	1	...	
	Stock (l)	0	0	0	0	0	0	0	0	...	

Figure 5. A. Grigsby et al. (2021) B. Ours

Attention-based LOB Highlighter & Forecaster

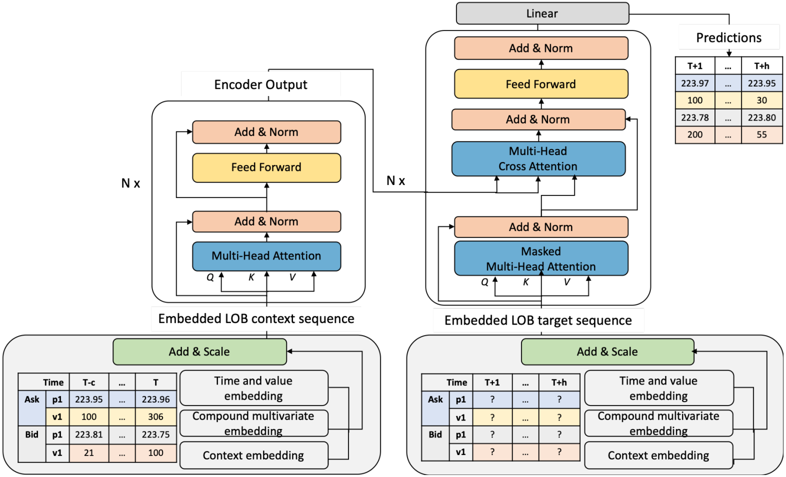


Figure 6. Attention-based LOB predictor network

Training Multi-level LOB Forecaster I

Percent change transformation for multi-level prices to enhance stationarity

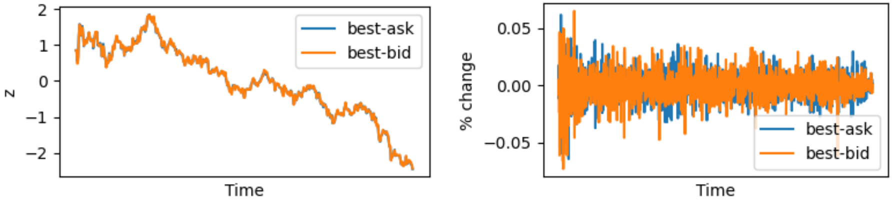


Figure 7. Value transformation methods: normalization (left) and percent change (right).

Min-max scale transformation is applied for multidimensional training.

Training Multi-level LOB Forecaster II

Multi-level LOB structural regularizer for ordinal penalty

For each *i*-th LOB snapshot, the multi-level price structure is given by:

$$p^a_{k_1 i} < p^a_{k_2 i} \quad \text{for all } k_1 < k_2, \quad p^b_{k_1 i} > p^b_{k_2 i} \quad \text{for all } k_1 < k_2$$
$$p^b_{k_1 i} < p^a_{k_2 i} \quad \text{for } k_1 = k_2 = 1,$$

We add some penalty if the multi-level ordinal structure is violated:

$$\begin{aligned}\text{Structure Loss}_i = \sum_{k=1}^{K-1} \text{ReLU}(\hat{p}^a_{k,i} - \hat{p}^a_{k+1,i}) + \text{ReLU}(\hat{p}^b_{1,i} - \hat{p}^a_{1,i}) \\ + \sum_{k=1}^{K-1} \text{ReLU}(\hat{p}^b_{k+1,i} - \hat{p}^b_{k,i}),\end{aligned}$$

where $p^j_{k,i}$ refers to the actual price, $\hat{p}^j_{k,i}$ refers to the predicted price.

Thus, $\text{Loss} = \text{Forecasting Loss (mse)} + w_o \cdot \sum_{i=1}^{I} \text{Structure Loss}_i$

Preliminary results

Summary of performance

	method (embedding)	Informer (temporal)	Spacetimeformer (spatiotemporal)	ours (compound multivariate)
mid-price	MSE	0.0177	0.0192	0.0124
	MAE	0.2187	0.2467	0.1456
price	MSE	0.0037	0.0039	0.0026
	MAE	0.0443	0.0490	0.0300
volume	MSE	0.0150	0.0142	0.0103
	MAE	0.0629	0.0602	0.0525
entire	MSE	0.0094	0.0090	0.0064
	MAE	0.0536	0.0546	0.0413
	structure loss	1.0791	0.7028	0.2928
Total loss (MSE + w*structure loss)		0.0201	0.0161	0.0094

Table 1. Limit order book prediction results. All values are scale-transformed values. We fix $w = 0.01$ to control the amount of penalty imposed by structure loss.

Training process

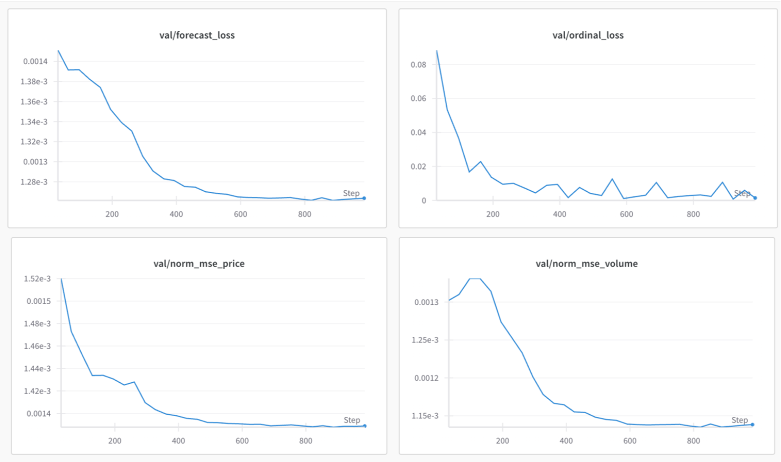


Figure 8. Training progress: validation loss outputs of entire error (upper left), structure loss (upper right), price error (bottom left), and volume error (bottom right).

Forecasting samples I

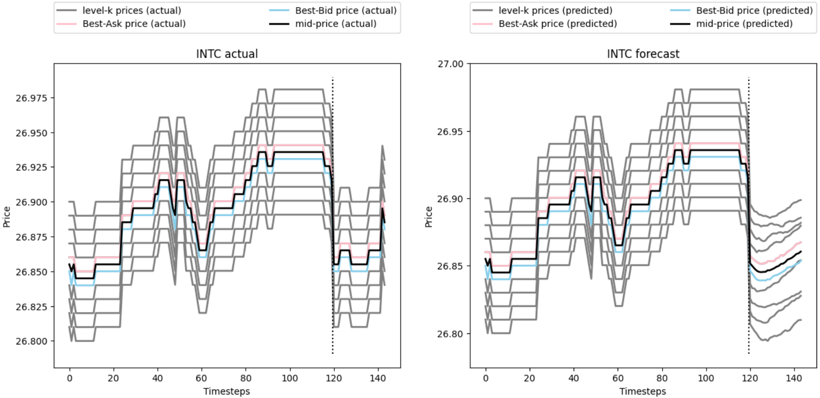


Figure 9. INTC level-5 price prediction: actual (left) and prediction (right)

Forecasting samples II

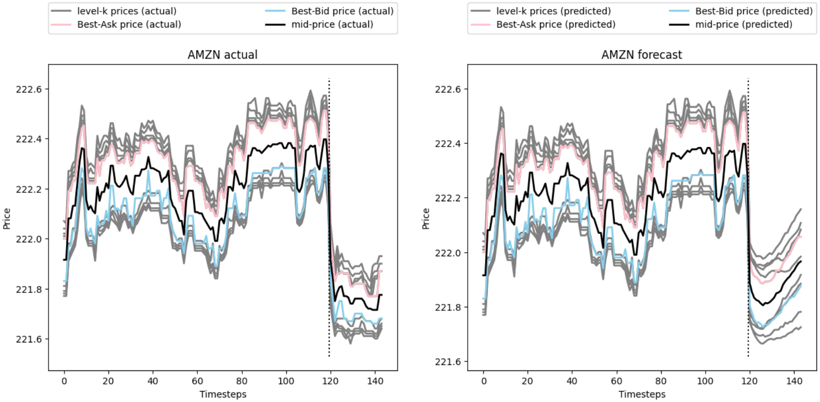


Figure 10. AMZN level-5 price prediction: actual (left) and prediction (right)

Forecasting samples III

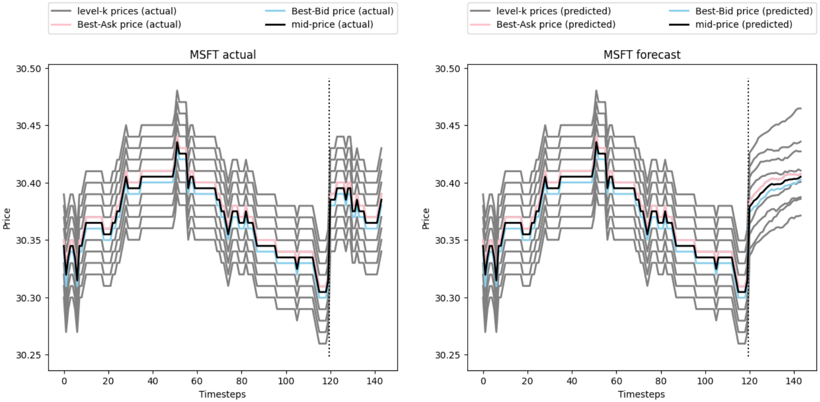


Figure 11. MSFT level-5 price prediction: actual (left) and prediction (right)

Forecasting samples IV

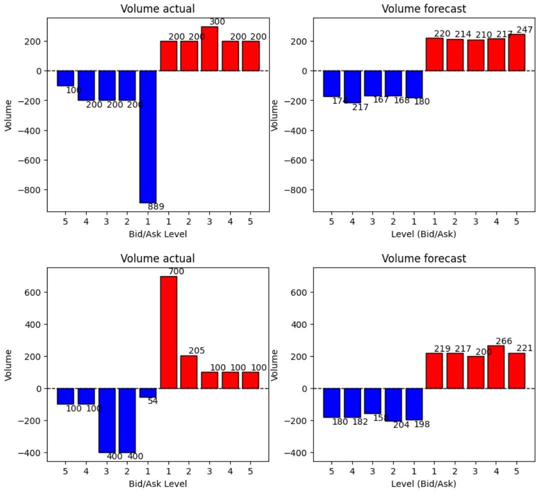


Figure 12. level-5 volume prediction samples: actual (left) and prediction (right)

Highlighter samples: self-attention matrices

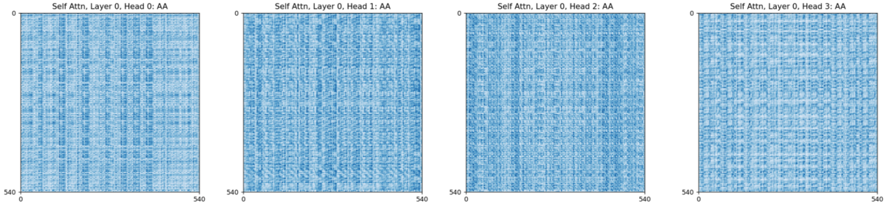


Figure 13. Self-attention four head samples.

Discussions and Future Directions

Discussions and Conclusions

- We trained our model on a subset of publicly available limit order book (LOB) data to evaluate its performance, and the results were promising.
- The limit order book (LOB) is a unique type of multivariate time series, characterized by its compound multivariate nature and multi-level ordinal structure. Our embedding process captured these features, and we incorporated structural restrictions by regularizing the loss function.
- Training the model on large volumes of high-frequency data requires significant computational resources. To apply our approach to the entire dataset, we must improve computational efficiency.

References I

- Fernandez-Tapia, J. (Sept. 2015) "Modeling, optimization and estimation for the on-line control of trading algorithms in limit-order markets". PhD thesis.
- Grigsby, J., Wang, Z., Nguyen, N., and Qi, Y. (2021) Long-range transformers for dynamic spatiotemporal forecasting. *arXiv preprint arXiv:2109.12218*.
- Kercheval, A. N. and Zhang, Y. (2015) Modelling high-frequency limit order book dynamics with support vector machines. *Quantitative Finance*, **15**(8), 1315–1329.
- Nevmyvaka, Y., Feng, Y., and Kearns, M. (2006) "Reinforcement learning for optimized trade execution". In: *Proceedings of the 23rd international conference on Machine learning*, pp. 673–680.
- Ntakaris, A., Magris, M., Kanniainen, J., Gabbouj, M., and Iosifidis, A. (2018) Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. *Journal of Forecasting*, **37**(8), 852–866.
- Park, B. and Van Roy, B. (2015) Adaptive execution: exploration and learning of price impact. *Operations Research*, **63**(5), 1058–1076.
- Sirignano, J. A. (2019) Deep learning for limit order books. *Quantitative Finance*, **19**(4), 549–570.

References II

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser,., and Polosukhin, I. (2017) Attention is all you need. *Advances in neural information processing systems*, **30**.

Zhang, Z., Zohren, S., and Roberts, S. (2019) Deeplob: deep convolutional neural networks for limit order books. *IEEE Transactions on Signal Processing*, **67**(11), 3001–3012.

Questions & Answers

- Contact information: jung320@purdue.edu
- Personal webpage: `https://jiwon-jung.github.io/`