Interim Report

COMP 4801

Final Year Project

Market Making with Deep Reinforcement Learning

Student: Aidan Wong Weng Seng
UID: 3035868918

Supervisor:

Prof. Huang Zhiyi

Abstract

Market making is recognized as a critical function in financial markets, providing liquidity and facilitating smoother trading through the continuous buying and selling of securities. Traditionally, market making has relied on rule-based models grounded in strong market assumptions, such as the Glosten and Milgrom model [1], and more recently, the widely respected Avellaneda and Stoikov framework [2]. With the rapid advancement of artificial intelligence and machine learning, particularly deep learning, new opportunities have emerged for transforming financial decision-making. Techniques such as Long Short-Term Memory (LSTM) networks [3] and Q-learning [4], when combined with order book dynamics, are offered as powerful tools for modeling complex market behavior. In this project, the stochastic control problem of market making is addressed through the development of a deep reinforcement learning agent capable of capturing intricate patterns in the limit order book. This objective is achieved through the development of two main components: the environment in which the agent is trained and the agent itself. The environment of the limit order book is modeled using a Hawkes process and constructed with the OpenAI Gym library. The agent is developed utilizing deep Q-learning (DQN) and is enabled with an LSTM that incorporates an attention mechanism. Currently, the environment has been successfully completed by mathematically modeling the arrival of orders in the limit order book, and the environment has been coded. The subsequent steps will involve the development of the agent using DQN and LSTM with an attention mechanism in PyTorch, followed by a comprehensive evaluation of the agent's performance compared to traditional market-making methods.

Acknowledgement

Gratitude is sincerely expressed to my brother and sister, Bryan and Chloe, for their tremendous assistance and support throughout this project. Their unwavering encouragement has been vital to the successful completion of this work.

Additionally, appreciation is extended to my supervisor, Dr. Huang, for his efforts and guidance during this project. His comments and insights have proven invaluable in the development and refinement of the research.

Table of Contents

List of Figures

List of Tables

Abbreviations

MM	Market Making
LOB	Limit Order Book
RL	Reinforcement Learning
DRL	Deep Reinforcement Learning
MDP	Markov Decision Processes
DNN	Deep Neural Networks
LLM	Large Language Models
DDPG	Deep Deterministic Policy Gradient
LSTM	Long Short-Term Memory
DQN	Deep Q-Network
RNN	recurrent neural networks
PnL	Profit and Loss
ML	Machine Learning

1 Introduction

1.1 Project Background

Market making is a foundational function in financial markets, essential for maintaining liquidity and ensuring the continuous flow of trades. Market makers actively quote both buy and sell prices for financial instruments, thereby reducing transaction costs and enabling smoother execution for other market participants. By standing ready to transact at publicly posted prices, they help narrow bid-ask spreads, absorb temporary imbalances in supply and demand, and stabilize price movements.

Market makers typically generate profits through two primary mechanisms: capturing the bid-ask spread and engaging in high-frequency trading strategies. The spread represents the difference between the prices at which they buy and sell securities, and its efficient management is central to their profitability. In modern electronic markets, many market makers operate algorithmically, using sophisticated models to adjust quotes dynamically in response to real-time market conditions, order flow, and inventory levels.

Their role becomes especially critical during periods of market stress or volatility, where their ability to provide liquidity can prevent sharp price dislocation and support overall market resilience. A key element that Market makers utilize to gain understanding of financial market condition is the Limit Order Book (LOB).

The Limit Order Book is a fundamental component of financial markets. It records all outstanding buy (bid) and sell (ask) orders that have not yet been executed, offering a real-time snapshot of market activity across various price levels and depths. Researchers study the Limit Order Book extensively to understand price action, liquidity, and market microstructure. By analyzing order flow, queue dynamics, and trade execution patterns, they uncover how prices evolve in response to supply and demand imbalances.

For example, Hasbrouck [10] provides a comprehensive framework for interpreting Limit Order Book data and its role in price formation. Cont., Stoikov, and Talreja [9] developed a stochastic model that captures empirical properties of high-frequency trading behavior. Xie et al. [11] applied Markov queue theory to model the Limit Order Book and optimize market maker strategies in China's A-share market. These studies demonstrate how the Limit Order Book serves as a rich source of information for modelling short-term price movements, estimating volatility, and designing trading algorithms.

Among the many emerging technologies in today’s age. One can leverage the capability of Deep Reinforcement Learning (DRL) to help better understand the LOB dynamics. DRL is the merger of two powerful techniques. The first is Reinforcement Learning (RL), a trial-and-error-based approach for modelling and making sequential decisions under uncertainty. It has been

widely used to solve stochastic optimal control problems framed as a Markov Decision Processes (MDPs).

The second is Deep Neural Networks (DNNs), which play a crucial role in today's world — especially with the rise of Large Language Models (LLM) that have significantly influenced modern society. These networks, composed of multiple layers, can learn complex patterns and representations from data, allowing them to approximate functions that map inputs to outputs with high accuracy.

With DRL, many achievements have been seen, such as mastering the ancient board game of Go through AlphaGo's neural network and tree search architecture [4], optimizing multi-agent traffic navigation via crowdsourced hyperparameter tuning in the Deep Traffic project [5], and enabling continuous control in robotic systems using the Deep Deterministic Policy Gradient (DDPG) algorithm [6].

Acknowledging these advances in DRL, the research aims to develop a deep reinforcement learning-based market-making model that leverages Long Short-Term Memory (LSTM) networks within a Deep Q-Network (DQN) framework. The model is designed to observe and learn temporal patterns in limit order book dynamics, enabling the agent to capture complex market behaviors and execute optimal quoting strategies.

Current academic progress of machine learning in the financial industry has seen significant growth across various domains. Oyewola, Akinwunmi, and Omotehinwa employed Deep LSTM Q-learning to analyze price movements in the oil and gas sector [12]. Wang explored the integration of deep learning techniques to estimate fill probabilities within a Limit Order Book (LOB) [13]. In the area of optimal market making, several researchers have contributed notable advancements. Lim and Gorse claim to be pioneers in this field, having constructed a market-making agent trained using the LOB framework proposed by Cont., Stoikov, and Talreja [9], as detailed in their reinforcement learning study [14]. Gasperov and Kostanjčar extended this work by incorporating a more realistic LOB model based on Hawkes processes [15]. Kumar proposed a novel deviation by constructing a real-time exchange simulation of the LOB, moving beyond traditional statistical modelling [16].

Despite these contributions, no existing research has yet combined a Hawkes-based LOB model with a deep reinforcement learning (DRL) framework that utilizes LSTM architecture. This research paper aims to explore such integration and evaluate its performance against traditional market-making models, such as the one proposed by Avellaneda and Stoikov [2].

1.2 Project Motivation

With the rapid advancement of artificial intelligence and machine learning, deep learning has emerged as a transformative tool in the financial sector. This research aims to leverage these technological advancements to address the shortcomings of conventional market-making strategies, which rely on rigid rule-based systems and fixed market conditions—often criticized

for their simplistic and overly optimistic assumptions. The objective is to enhance trading strategies, improve quoting accuracy, and optimize inventory risk management.

A key breakthrough in this area was demonstrated by Lim and Gorse, who showed that reinforcement learning using a simple discrete Q-learning algorithm outperformed the widely respected Avellaneda and Stoikov model, which has long served as a benchmark in the field [14]. Building on this, Oyewola, Akinwunmi, and Omotehinwa introduced Long Short-Term Memory (LSTM) networks into a DRL framework, yielding promising results in capturing temporal dependencies in financial data. These developments have paved the way for modelling the Limit Order Book (LOB) using Hawkes processes to train DRL agents that leverage LSTM architectures. This approach aims to capture time-dependent patterns in LOB dynamics and address the challenge of optimal market making more effectively.

Through this research, it is hoped that the adoption of more advanced machine learning techniques will be inspired by demonstrating their potential applications within the financial industry.

1.3 Project Objective

The primary objective of this research is to design and implement an intelligent agent-based market maker using deep reinforcement learning techniques, capable of interpreting and reacting to the complex dynamics of a financial limit order book. The agent will be trained to analyze key market microstructure features such as order book dynamics. These elements will be combined with strategic decision-making frameworks like the Kelly criterion to guide the agent's quoting behavior. The overarching goal is to build a robust market-making model that leverages Long Short-Term Memory (LSTM) networks to capture temporal dependencies in order book movements, while employing a Deep Q-Network (DQN) to learn optimal pricing strategies. The agent will aim to maximize profitability, maintain quoting precision, manage inventory efficiently, and mitigate exposure to market risk.

1.4 Project Scope and Deliverables

The primary deliverable of this study is a comprehensive research report that compares a modernized market-making approach—powered by deep reinforcement learning (DRL)—with traditional models. To achieve this, the project is organized around six key milestones: first, a literature review on stochastic models for the limit order book (LOB) will be conducted; second, parameter estimation for the LOB Hawkes model will take place. Next, a simulated training environment will be developed, followed by the development of the agent and research on the Q-function. The project will then include benchmarking against traditional market-making models, culminating in the final report of results.

1.5 Report Outline

Section 2 outlines the methodology, detailing the models, tools, and rationale behind the experimental setup; Section 3 presents interim findings and expected results based on the agent's performance and benchmarking; and Section 4 concludes the study with a summary of insights and suggestions for future improvements.

2 Project Methodology

In the realm of reinforcement learning, two fundamental components are essential: the environment and the agent. This section will elucidate the specific characteristics and rationale underlying the selection of the chosen environment.

2.1 Environment

The construction of the agent simulation environment for repetitive training will be modeled based on the Hawkes Process theory proposed by Lu and Abergel [18]. This model was selected because it effectively replicates real-world continuous order book dynamics by capturing short-term dependencies and changes in the limit order book, unlike the model by Cont., Stoikov, and Talreja [9], which does not account for time-dependent properties of the financial market. Consequently, in comparison to traditional stochastic limit order book models, the Hawkes Process model provides a more accurate representation of the real-time limit order book by capturing both excitation and inhibition effects among different types of orders, thus offering a more realistic depiction of market behavior.

As shown in Figure 1 below. Reproduced from [18], this diagram highlights the significance of employing Hawkes' process in modelling the order book. Shown in green and red, the Hawkes process matches closely with the theoretical order book reaction in blue. Whereas the independent Poisson process deviates more strongly from the theoretical model. This result serves as a strong justification to model the order book using the theory proposed by Lu and Abergel [18].

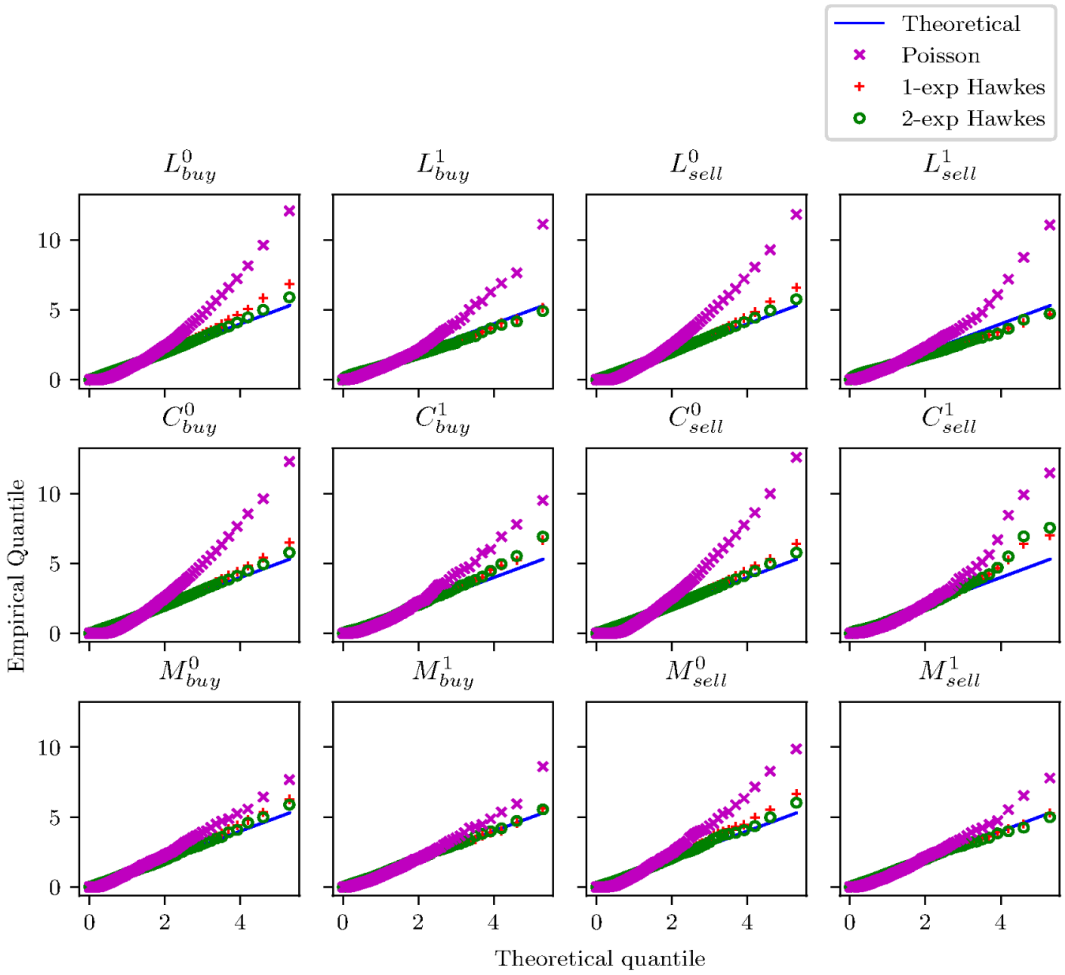

Figure 1: Q-Q plot goodness of fit tests for order book. [18]

To code the environment, OpenAI gym would be utilised. This is an open-source python module which enables researchers and data scientists to build a reinforcement learning environment with ease. OpenAI gym provides easily integrated custom environments that allow benchmarking across a wide range of tasks and comparing performance using consistent metrics. This enables us to accurately and consistently compare different agent models and market-making strategies, which is essential for validating new approaches.

2.2 Agent

To construct a reliable and effective model, it is imperative to incorporate deep learning networks into our framework. These additional layers of networks possess the capability to learn intricate patterns from order book dynamics, thereby enabling the agent to identify hidden temporal dependencies between each action.

The agent is built using LSTM layers along with a Q-function to capture long and short-term dependencies that are present in the training data. LSTM layers offer the agent the capability

of maintaining a memory of past observations, providing more knowledge through each decision-making process. This further enhances the agent's capability of spotting hidden temporal relationships between different past and present states.

Financial markets are widely argued to contain high levels of noisy data, often influencing the models' decision-making and hypothesis. Hence, with this issue, introducing LSTM with an Attention Mechanism improves performance by allowing the network to focus on the most relevant parts of the input sequence. Proposed by Sang and Li [8], this model has yielded positive results in modelling temporal dependencies in financial markets. As shown in Figure 2a and Figure 2b below by Sang and Li [8] revealed that an LSTM with an attention mechanism outperforms the normal LSTM in capturing the predicted closing price.

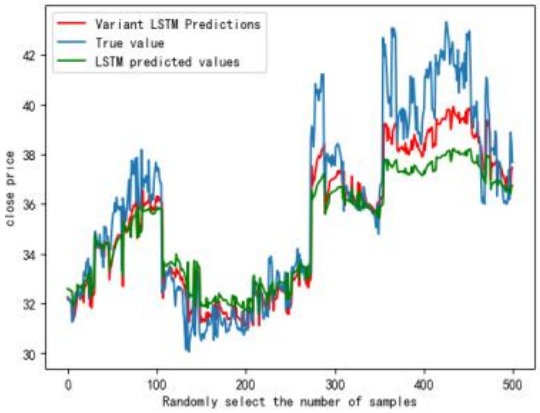

Figure 2a: LSTM [8]

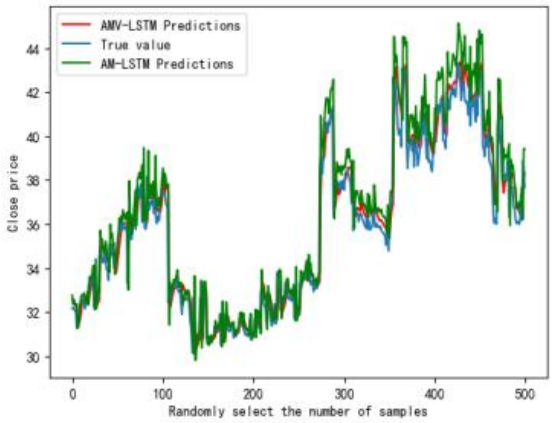

Figure 2b: LSTM with Attention [8]

When modelling the Q-function, which serves as the reward function, the agent considers factors such as profit and loss (PnL), inventory risk, and fill rates to guide optimal quoting decisions. One notable idea, implemented by [14], is the use of the Constant Absolute Risk Aversion (CARA) utility function to shape the reward structure. This approach introduces risk sensitivity into the agent's decision-making process and serves as the foundation for constructing the policy that drives optimal market-making behavior.

3 Current Progress

As outlined in the methodology section, there are two primary components in reinforcement learning research that are essential: the environment in which the agent trains and the agent itself. In the following section, the development of the agent will be elaborated upon as part of the current progress.

3.1 Environment Development

Following the completion of the initial literature review, attention is directed toward the development of the environment. Given the financial structure of a limit order book, accurate modeling of this system is essential for replicating the real-world financial dynamics present in the equities market.

To facilitate this, the Hawkes process has been employed to mathematically model the arrival of orders in a mutually dependent manner. Subsequently, the environment was coded, as illustrated in Figure 3 below.

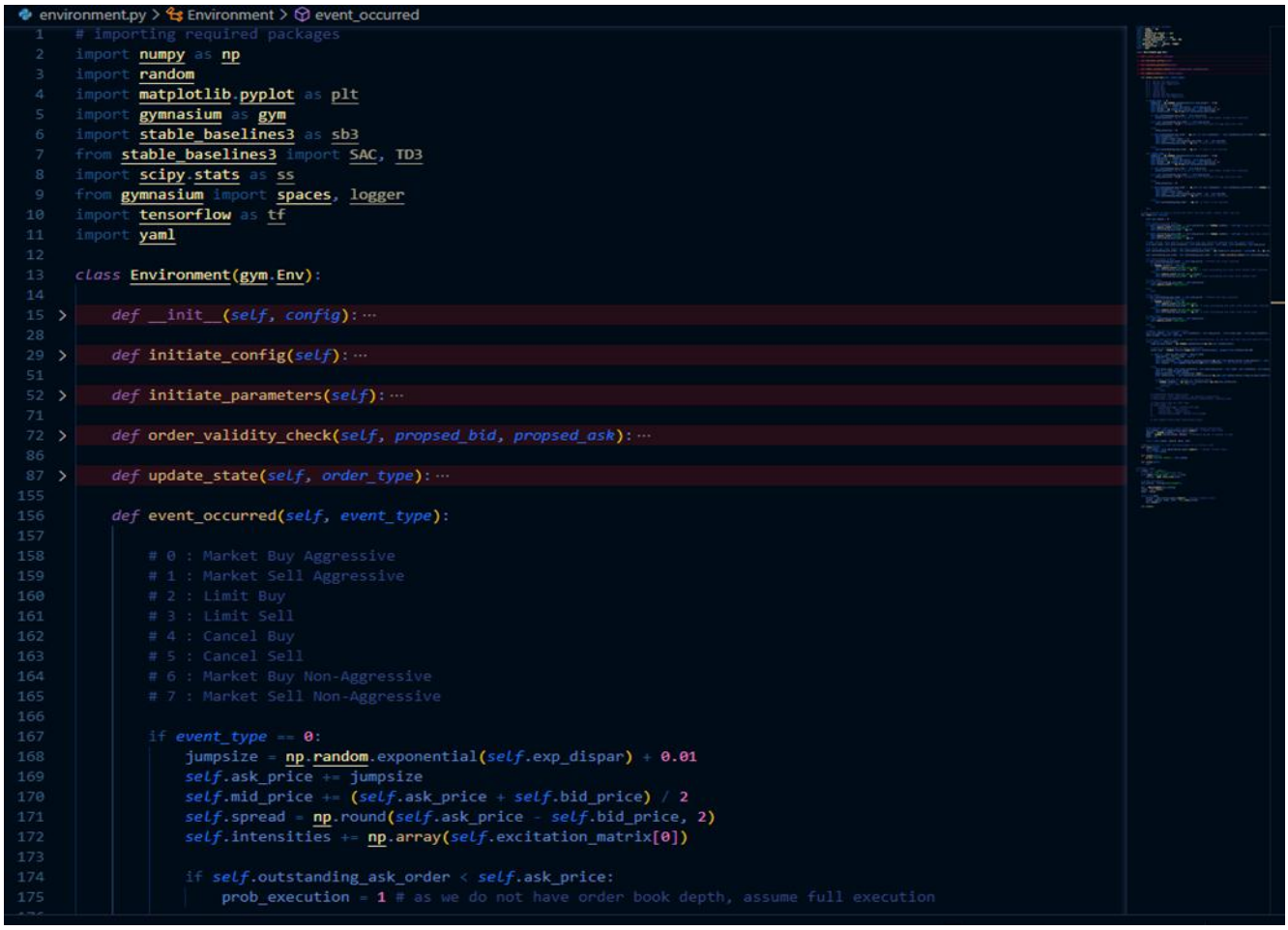

Figure 3: Code of Environment

As demonstrated above, the current progress in mathematically modeling the order arrivals within the limit order book has been achieved. Furthermore, the code for the environment has been completed using the OpenAI Gym library.

3.2 Difficulties encountered

Modeling the environment represents a challenging endeavor, particularly due to the necessity of ensuring that a flawed or erroneous environment is not constructed, as such a misstep could distort return results. Establishing that the mathematical model meets the required standards of robustness is essential, thereby enabling effective training of the agent and the production of accurate outcomes. As noted by Law and Viens [19], much of the current research on market-making approaches remains inconsistent with respect to direction, timing, and volume, which can result in illusory gains in back testing results.

3.3 Mitigations

Utilizing the research conducted by Law and Viens [19], a weakly consistent limit order book model has been developed, capturing the interdependent dynamics between price fluctuations and limit order book behavior, specifically with respect to direction and timing.

The erroneous assumption in the original Avellaneda and Stoikov model [2], which posits that price movements are independent of the arrival of market orders and limit order book dynamics, has been rectified in the weakly consistent pure jump market model proposed by Law and Viens [19]. Consequently, the environment has been constructed based on the proposed model by Law and Viens [19].

3.4 Remaining work plan and Proposed Schedule

At present, of the two sections—Environment and Agent—the environment component has been completed, and efforts are focused on finalizing the agent section within the remaining timeframe. As detailed in Table 1 below, commencing from September 2025, the literature review has been completed, formulating the ideas and concepts of this research. Various sources of research have been analyzed and studied to gain inspiration and assess the feasibility of the project.

Beginning from week 5, thorough mathematical modeling of the limit order book has been conducted, including a study of the Hawkes process and its application to this research. Progress has also been made in coding the environment using the OpenAI Gymnasium library, where multiple scenarios and types of order events have been implemented.

In the upcoming weeks, the focus will shift to agent development (weeks 13 to 24), during which the agent will be built using DQN and LSTM with an attention mechanism in PyTorch, followed by model evaluation (weeks 25 to 29) of the results. After completing the necessary benchmarking, the research will be concluded in the final weeks (weeks 30 to 32), wrapping up insights from the final year project.

Phase	Timeline	Status	Key Deliverables
Literature Review	1~4 weeks	Completed	Develop Theoretical Framework
LOB Simulation	5~12 weeks	Completed	Functional Environment with trainable model
Agent Development	13~24 weeks	Planned	Effective Q-function and agent decisions
Model Evaluation	25~29 weeks	Planned	Traditional agents and Benchmark analysis
Conclusion	30~32 weeks	Planned	Final Year Report

Table 1: Projected timeline, status and key deliverables

4 Conclusion

In conclusion, the objective of this research is to integrate the fields of market making and deep reinforcement learning, with the hope of facilitating intelligent market-making through machine learning. This study challenges traditional rule-based methods of market making by employing deep reinforcement learning techniques.

Key achievements during the current progress include the mathematical modeling of limit order book dynamics, specifically relating to price movements and the arrival of orders, utilizing jump processes and Hawkes processes, respectively. Additionally, the validity of the mathematical environment has been assessed, and the development of the environment using the OpenAI Gym library has been successfully completed.

The significance of this work lies in the consistent mathematical modeling of the limit order book, which avoids errors in back testing results while maintaining simplicity in design. Furthermore, developing the environment to align with real-life trading scenarios represents an important outcome. The code accommodates various conditions, including the arrival of up to eight different mutually dependent orders, price jumps, inventory management, and market fees associated with real-world exchanges.

Although the model proposed by Law and Viens [19] demonstrates the ability to yield reliable results, there is still room for improvement, as it does not account for the depth of orders beyond the best bid and ask. This limitation could hinder a comprehensive study of the dynamics within the limit order book environment.

Moreover, the agent operates within a zero-sum game framework, which reflects real-world scenarios where multiple market makers compete to achieve optimal quotations and profit.

Thus, the research is constrained in this regard, particularly when considering multiple participants, where other agents may seek to minimize their gains.

Acknowledging these limitations opens avenues for future research. Developing more advanced models of the limit order book that fully account for depth would enhance our understanding of market mechanics.

Additionally, robust adversarial reinforcement learning could be investigated, as it offers a means for greater generalization and robustness under conditions of model uncertainty. Incorporating these enhancements could bring us closer to accurately mimicking real-world market dynamics and improving agent capabilities in market making.

References

[1] L. R. Glosten and P. R. Milgrom, “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,” *Journal of Financial Economics*, vol. 14, no. 1, pp. 71–100, Mar. 1985. [Online]. Available: https://milgrom.people.stanford.edu/wp-content/uploads/1984/09/Bid-Ask-and-Transaction-Prices.pdf

[2] M. Avellaneda and S. Stoikov, “High-frequency trading in a limit order book,” *Quantitative Finance*, vol. 8, no. 3, pp. 217–224, Apr. 2008. [Online]. Available: https://people.orie.cornell.edu/sfs33/LimitOrderBook.pdf

[3] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6795963

[4] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, May. 1992. [Online]. Available: https://link.springer.com/article/10.1007/BF00992698

[5] D. Silver et al., “Mastering the game of Go with deep neural networks and tree search,” *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016. [Online]. Available: https://www.nature.com/articles/nature16961

[6] L. Fridman et al., “DeepTraffic: Crowdsourced hyperparameter tuning of deep reinforcement learning systems for multi-agent dense traffic navigation,” *arXiv preprint*, arXiv:1801.02805, Jan. 2018. [Online]. Available: https://arxiv.org/abs/1801.02805

[7] S. Sang and L. Li, “A novel variant of LSTM stock prediction method incorporating attention mechanism,” *Mathematics*, vol. 12, no. 7, p. 945, Mar. 2024. [Online]. Available: https://doi.org/10.3390/math12070945

[8] M. Hausknecht and P. Stone, “Deep recurrent Q-learning for partially observable MDPs,” *arXiv preprint*, arXiv:1507.06527, Jul. 2015. [Online]. Available: https://arxiv.org/abs/1507.06527

[9] R. Cont, S. Stoikov, and R. Talreja, “A Stochastic Model for Order Book Dynamics” *SSRN Electronic Journal*, Sep. 2008. [Online]. Available:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1273160

[10] J. Hasbrouck, “Limit Order Market,” in *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*, Oxford University Press, Jan. 2007. [Online]. Available: https://books.google.com.hk/books?id=aaReNv846eMC

[11] F. Xie, Y. Liu, C. Hu, and S. Liang, “Dynamic modeling of limit order book and market maker strategy optimization,” *Mathematics*, vol. 13, no. 5, Feb. 2025. [Online]. Available: https://www.mdpi.com/2227-7390/13/5/778

[12] A. Oyewola, O. Akinwunmi, and O. Omotehinwa, “Deep LSTM Q-learning for price movement prediction in the oil and gas sector,” *Knowledge-Based Systems*, vol. 281, Jan. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0950705123010389

[13] C. Maglaras, C. C. Moallemi, and M. Wang, "A deep learning approach to estimating fill

probabilities in a limit order book," Quantitative Finance, vol. 22, no. 11, pp. 1989-2003, Nov. 2022. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/14697688.2022.2124189

[14] Y.-S. Lim and D. Gorse, "Reinforcement learning for market making in a limit order book," presented at the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018), Apr. 2018. [Online]. Available: https://www.esann.org/sites/default/files/proceedings/legacy/es2018-50.pdf

[15] M. Gasperov and S. Kostanjčar, "Market making with Hawkes process limit orderbook," *arXiv preprint*, arXiv:2207.09951, Jul. 2022. [Online]. Available: https://arxiv.org/abs/2207.09951

[16] P. Kumar, "Deep Hawkes Process for High-Frequency Market Making," *arXiv preprint*, arXiv:2109.15110, Sep. 2021. [Online]. Available: https://arxiv.org/abs/2109.15110

[17] J. Spooner, D. Savani, and J. Zohren, "Deep reinforcement learning for market making in high-frequency trading," *HAL open archive*, Jan. 2018. [Online]. Available: https://hal.science/hal-01686122v1/document

[18] X. Lu and F. Abergel, "High dimensional Hawkes processes for limit order books: Modelling, empirical analysis and numerical calibration," *Quantitative Finance*, vol. 18, no. 8, pp. 1315–1330, Jan. 2018. [Online]. Available: https://hal.science/hal-01686122v1/document

[19] Law, B and Viens, F. "Market Making under a Weakly Consistent Limit Order Book," *High Frequency,* vol. 2, no.4, pp. 215-238, Aug. 2019 [Online]. Available: https://arxiv.org/pdf/1903.07222