Detailed Project Plan

Market Making with Deep Reinforcement Learning

Student: Aidan Wong Weng Seng
UID: 3035868918
Supervisor: Dr. Huang Zhiyi
Project code: fyp25009

1 Project Background

1.1 Abstract

Market making is a critical function in financial markets, providing liquidity and facilitating smoother trading by continuously buying and selling 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, offer powerful tools for modeling complex market behavior. In this project, we aim to tackle the stochastic control problem of market making by developing a deep reinforcement learning agent capable of capturing intricate patterns in the limit order book.

1.2 Market Making

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 dislocations and support overall market resilience.

1.3 Traditional Market Making Models

In this section, we examine two foundational market-making models that operate under distinct assumptions about market structure and participant behavior. These models have shaped the theoretical underpinnings of liquidity provision and continue to influence modern algorithmic trading strategies

1.3.1 Glosten and Milgrom Model

Introduced by Lawrence Glosten and Paul Milgrom in 1985 [1], this model addresses the challenge of optimal price quoting in markets characterized by asymmetric information. It assumes the presence of two types of traders: informed traders, who possess private knowledge about the true value of the asset, and uninformed traders, who trade for liquidity or other non-informational reasons.

The market maker, lacking knowledge of the counterparty's type, must continuously adjust bid and ask quotes to mitigate the risk of adverse selection. This is achieved by updating beliefs about the asset's value based on the direction and frequency of incoming orders. For instance, a sequence of buy orders may signal informed trading, prompting the market maker to raise the ask price. The resulting bid-ask spread reflects the compensation required for bearing informational risk, and its width increases with the perceived likelihood of informed trading.

This model elegantly explains why spreads persist even in competitive markets and laid the groundwork for subsequent microstructure theory.

1.3.2 Avellaneda-Stoikov Model

Proposed by Marco Avellaneda and Sasha Stoikov in 2008 [2], this model offers a quantitative framework for optimal market-making within a limit order book. Unlike Glosten and Milgrom's information-based approach, Avellaneda-Stoikov focuses on inventory management and risk control over a finite time horizon.

The model formulates the market-making problem as a stochastic control task, where the agent seeks to maximize expected utility of terminal wealth while managing inventory risk. It derives a system of ordinary differential equations from the Hamilton-Jacobi-Bellman (HJB) equation, leading to closed-form expressions for optimal bid and ask quotes. These quotes are dynamically adjusted based on the agent's current inventory, market volatility, and order arrival intensity.

1.4 Deep Reinforcement Learning Approach

With the rapid advancement of technology, particularly in artificial intelligence, deep reinforcement learning has emerged as a transformative tool in this domain. Unlike traditional market making, which is rule-based on predefined market conditions that are often criticized for containing strong, naïve assumptions, this research aims to tackle these limitations by exploring the intersection of deep reinforcement learning and market making. The goal is to enhance trading strategies, improve quoting accuracy, and optimize inventory risk management.

This 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.

1.4.1 Deep Reinforcement Learning

Deep reinforcement learning (DRL) is the merger of two powerful techniques. The first is Reinforcement Learning (RL), a trial-and-error-based approach for modeling and making sequential decisions under uncertainty. It has been widely used to solve stochastic optimal control problems framed as 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 (LLMs) that have significantly influenced modern society. These networks, composed of multiple layers, are capable of learning complex patterns and representations from data, allowing them to approximate functions that map inputs to outputs with high accuracy.

With Deep Reinforcement Learning (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 DeepTraffic project [5], and enabling continuous control in robotic systems using the Deep Deterministic Policy Gradient (DDPG) algorithm [6].

1.4.3 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a specialized form of recurrent neural networks (RNNs) designed to overcome the limitations of traditional RNNs in learning long-term dependencies. Introduced by Hochreiter and Schmidhuber in 1997 [7], LSTMs use a gated architecture — including input, output, and forget gates — to regulate the flow of information and selectively retain or discard memory over time. This structure enables LSTMs to effectively model sequential data with temporal dependencies, making them particularly useful in domains such as natural language processing, time-series forecasting, and financial modeling. In reinforcement learning, LSTMs are often integrated into agents to help maintain a memory of past observations, allowing for more informed decision-making in partially observable environments [8].

1.4.4 Stochastic Model of LOB

In order for the agent to continuously update its reward and penalty functions, we need to develop a simulation environment where it can interact, learn, and refine its decision-making process over time.

To model the dynamic behavior of the limit order book in financial markets, we adopt the framework proposed by Rama Cont, Sasha Stoikov, and Rishi Talreja [9], which has become a widely used reference among researchers. Their model captures the stochastic nature of order book dynamics using Poisson processes, birth-and-death mechanisms, and power-law distributions to describe order arrivals, cancellations, and executions. This paper laid the groundwork for quantitative modeling of microstructure, influencing both academic research and industry applications in algorithmic trading, market making, and liquidity analysis.

2 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.

To achieve this, the project will be divided into a series of structured phases, each building upon the last to ensure a coherent and scalable development process. The initial phase focuses on constructing a simulated exchange environment grounded in the stochastic order book dynamics model proposed by Rama Cont, Sasha Stoikov, and Rishi Talreja [9]. This model offers a mathematically rigorous framework for capturing the evolution of limit order

books and will serve as the foundation for the agent's learning environment. A thorough review and understanding of Cont's methodology will precede its replication and implementation in code, resulting in a controlled simulation where the agent can interact, adapt, and refine its decision-making strategies.

Once the environment is established, the next step is to equip it with analytical tools that allow the agent to extract meaningful features from the order book. These tools will compute relevant metrics—such as spread, depth, imbalance, and arrival rates—that inform the agent's decision-making process. With these inputs in place, a preliminary deep learning agent will be developed, integrating LSTM for sequential pattern recognition and Q-learning for reinforcement-based strategy optimization. This agent will begin interacting with the environment, learning from its experiences and refining its quoting behavior over time.

Following the agent's initial deployment, the Q-function will be modeled to define the reward structure that governs the agent's learning. This involves specifying the conditions under which the agent receives positive or negative feedback, based on factors such as profit generation, inventory levels, and market impact. Designing an effective reward mechanism is critical to shaping the agent's policy and ensuring that its actions align with the desired market-making objectives. This phase will require extensive experimentation and theoretical analysis to balance short-term gains with long-term stability.

Once the agent has been trained on a sufficiently large and diverse dataset, its performance will be evaluated through benchmark analysis. This involves comparing the agent's behavior and outcomes to those of traditional market-making models, such as the Glosten and Milgrom framework. By assessing metrics like profitability, quoting accuracy, and risk exposure, the study will highlight the strengths and limitations of applying deep learning techniques to market-making tasks.

The final deliverable of this research will be a fully functional market-making agent operating within a custom-built simulation. This agent will demonstrate the feasibility and potential of integrating deep reinforcement learning into financial market microstructure modeling, offering insights into how intelligent systems can enhance liquidity provision and price discovery in modern electronic markets.

3 Project Methodology & Schedule

3.1 Literature Review and Theoretical framework (Weeks 1 ~ 4)

- Understand and review stochastic order book dynamics model proposed by Rama Cont, Sasha Stoikov, and Rishi Talreja.
- Study how this project could be feasible and learn current research studies on similar topics.
- Identify and learn Technical resources or tools to be used to develop the agent mode.

3.2 Develop Limit Order Book Exchange Simulator (Weeks 5 ~ 12)

- Study and research on Deep Reinforcement Learning environment construction.
- Transform the mathematical models present in stochastic order book dynamics paper into code.
- Design environment for simulation such that the model is able to learn and interact with the environment.
- Build tools and functions so that the agent is able to get statistics from the limit order book to turn them into actionable items.
- Setup virtual machine and resource for model to run on the cloud.

3.3 Agent Development (Weeks 13 ~ 24)

- Research and development of reinforcement learning agents, create a minimal viable agent first.
- Research on improving Q-function to reward and penalise agent action accordingly. Could take inspiration from economical principles such as CARA utility.
- Consider other metrics to improve models, this includes and not limited to:
    - Risk management with respect to inventory
    - Quote creation
    - Fill rates
    - PnL
- Agent Training & Fine tuning:
    - Tackle questions such as the number of layers of LSTM to use to improve the model.
    - Additional metrics or hyper parameter tuning to improve model performance.
    - Research on decision making algorithmics that may be helpful (Optional)

3.4 Model Evaluation (Weeks 25 ~ 29)

- Construct traditional market making models:
    - Glosten and Milgrom Model
    - Avellaneda-Stoikov Model
- Benchmark analysis, compare model performance in terms of:
    - PnL
    - Inventory risk
    - Sharp Ratio

3.5 Conclusion (Weeks 30 ~ 32)

- Report of results of study
- Presentation and Poster for research study.
- Draw conclusion and effectiveness of using DRL.

Summary

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

4 Conclusion

This project aims to conduct a comprehensive research on the potential capabilities of employing deep reinforcement learning in achieving optimal market making quotation. Using the capabilities of LSTM, Q-learning, Deep Reinforcement Learning and order book dynamics we attempt to uncover the power of Artificial Intelligence in the financial markets.

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.

[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.

[3] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[4] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, May 1992.

[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.

[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.

[7] T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning,” *arXiv preprint*, arXiv:1509.02971, Sep. 2015.

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

[9] R. Cont, S. Stoikov, and R. Talreja, “The dynamics of order book: An empirical analysis of high frequency data,” *SSRN Electronic Journal*, Sep. 2008. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1273160