Replication for 'Market Making With Signals Through Deep Reinforcement Learning'

Jiaxing Wei
jwei2002@uw.edu

1 Paper Summary

1.1 Introduction

The project involves replicating the paper *Market Making With Signals Through Deep Reinforcement Learning* (Gasperov et al., 2021). This study introduces a model-free Deep Reinforcement Learning (DRL) framework designed to address the limitations of traditional stochastic control models (e.g., Avellaneda-Stoikov). The core innovation lies in the integration of standalone Signal Generating Units (SGUs) that provide high-frequency alpha predictions specifically price range and trend forecasts into the RL state space. The goal is to develop an agent that adaptively manages the bid-ask spread to maximize terminal wealth while significantly reducing inventory risk in volatile markets.

1.2 Main Body

- **State Space & Signal Integration**: The state space $S_t = [I_t, RR_t, TR_t]$ is three-dimensional, consisting of the agent's current inventory (I_t) and two predictive signals: the modified realized price range (RR_t) and the price trend (TR_t). Unlike traditional models, it intentionally excludes time remaining to suit continuous trading environments. These signals serve as gating mechanisms, enabling the agent to adjust its quoting strategy in response to market volatility and momentum forecasts.

- **Action Space & Continuous Offsets**: The framework utilizes a two-dimensional, continuous action space $A_t = [A_{t,1}, A_{t,2}]$. These actions represent bid and ask offsets relative to the current best bid/ask prices, rather than absolute prices. This formulation ensures the quoting is tick-based and inherently accounts for the prevailing market spread, facilitating both aggressive (inside-spread) and conservative quoting.

- **Reward Shaping & Risk Control**: The reward function R_{t+1} balances captured spreads with an absolute inventory penalty: $\lambda|I_{t+1}|$. This linear penalty, inspired by Value-at-Risk (VaR) interpretations, discourages large directional exposures. The reward is designed to disincentivize trend-chasing and focus the agent on round-trip spread capturing, which is crucial for maintaining market neutrality.

- **Neuroevolution & Adversarial Training**: To avoid the noisy gradient problem common in financial RL, the paper employs neuroevolution via genetic algorithms to optimize the policy. Furthermore, it introduces Adversarial Reinforcement Learning (ARL), where an adversary agent strategically perturbs the market maker's quotes. This minimax optimization approach enhances the model's robustness against model misspecification and changing market regimes.

1.3 Hypothesis Extraction

1. Object of Analysis: High-frequency Limit Order Book (LOB) data and HFT alpha signals of the China A-share CSI 300 ETF.
2. Dependent Variable: Inventory-adjusted quoting returns (Terminal Wealth) and risk metrics, including Maximum Drawdown (MDD) and Mean Absolute Position (MAP).
3. Independent Variable: DRL-generated tick offsets, SGU-provided price range/trend predictions, and current inventory levels.
4. Expected Outcome: The signal-gated DRL agent will outperform standard benchmarks (FOIC and GLFT) by achieving higher terminal wealth with lower inventory risk, demonstrated through a superior PnL-to-MAP ratio.
5. Verification Method: Walk-forward backtesting on large-scale historical datasets with strict data-leakage protocols and GPU-accelerated simulation across different volatility regimes.

1.4 Conclusion

Integrating HFT alpha signals with neuroevolutionary RL provides a robust solution for market making. The implementation demonstrates that signal-gating and optimized computational primitives effectively balance profitability and inventory risk in the China A-share market.

2 Literature Review

The problem of optimal Market Making (MM) is fundamentally a challenge of stochastic inventory control. The objective is to simultaneously quote bid and ask prices to capture the spread while managing the risk of holding a non-zero inventory.

2.1 Foundational Models and Reinforcement Learning

Traditional market making models, such as the seminal Avellaneda-Stoikov (AS) framework, rely on solving Hamilton-Jacobi-Bellman (HJB) equations under strict stochastic assumptions. However, these models often struggle with the non-linearities and high-dimensional state spaces of real-world Limit Order Books (LOB). Consequently, recent literature has shifted toward Reinforcement Learning (RL) to learn optimal policies directly from data.

Spooner et al. (2018) in 'Market Making via Reinforcement Learning' pioneered the use of temporal-difference learning (Sarsa) with tile coding. They focused on the importance of the reward function, demonstrating that purely symmetric PnL-based rewards lead to sub-optimal inventory management. This work established the baseline for using asymmetric risk aversion to control inventory skew.

2.2 Deep Reinforcement Learning and LOB Feature Extraction

With the advent of Deep RL, researchers began utilizing neural networks to process high-dimensional LOB data without manual feature engineering.

- Sadighian (2019) in 'Deep Reinforcement Learning in Cryptocurrency Market Making' advanced this by using Advanced Policy Gradient (A2C) and PPO algorithms. This study highlighted the effectiveness of Recurrent Neural Networks (RNNs) in capturing the temporal dependencies of order flow, particularly in volatile cryptocurrency markets.
- Guo et al. (2020/2023) in 'Market Making with Deep Reinforcement Learning from Limit Order Books' introduced an Attn-LOB architecture. By employing attention mechanisms, they demonstrated that an agent could selectively focus on the most informative levels of the LOB, significantly outperforming traditional CNN-based feature extractors.

2.3 Adversarial Robustness and Training Stability

A recurring issue in DRL for market making is the sim-to-real gap and the tendency of agents to overfit to specific market regimes.

- Spooner & Savani (2020) in 'Robust Market Making via Adversarial Reinforcement Learning' addressed this by framing the MM problem as a zero-sum game between the market maker and an adversarial disturber. This forced the agent to learn policies that remain profitable even under unfavorable price movements.
- Gasperov et al., (2021) build upon this adversarial foundation but introduce a critical innovation: Signal Generating Units (SGUs). Unlike previous end-to-end models, they decouple the prediction of market volatility (via XGBoost) and trend (via LSTM) from the RL decision-making process. Furthermore, they utilize Neuroevolution instead of gradient descent to avoid the instability of noisy financial gradients.

2.4 Research Gaps: A-Share ETF300 Constraints

Despite these advancements, the current body of literaturelargely focused on T+0 markets like Crypto or FX, which exhibits a significant gap when applied to the A-share ETF300 market.

1. **The T+1 Liquidity Gap**: Existing models (Spooner, Sadighian, and Gasperov) assume that inventory can be liquidated at any time. In the

A-share market, the T+1 rule creates an asymmetric liquidity trap. Assets bought during the current session cannot be used to hedge sell-side pressure until the following day.

2. **The Price Limit Constraint**: None of the referenced papers explicitly models the 10% price limit mechanism. In A-shares, as an ETF approaches a limit-up or limit-down state, the execution probability function (λ) becomes binary rather than decaying. This renders standard execution models, such as those used in Gasperov (2021), ineffective near the boundaries.