Building an Algorithmic Black Box

A Market Microstructure–Driven Reinforcement Learning Trading System

Kowshik

Roll No: 24B2492

Department of Metallurgical Engineering and Materials Science

January 28, 2026

Abstract

This project presents the design and evaluation of a fully custom algorithmic trading system constructed from first principles. The system integrates a realistic limit order book, an event-driven exchange simulator, heterogeneous trading agents, and a reinforcement learning (RL) trader trained using Proximal Policy Optimization (PPO).

Unlike traditional price-based backtesting approaches, this work emphasizes market microstructure realism by explicitly modeling order matching, liquidity provision, inventory risk, and asynchronous agent interaction. The trading environment is formalized as a Markov Decision Process (MDP) and implemented as a Gymnasium-compatible environment.

A sequence of controlled experiments is conducted to study emergent market behavior, including volatility clustering, fat-tailed returns, and herding dynamics. While the RL agent does not consistently outperform simple baselines in raw returns, it demonstrates improved risk-adjusted performance under specific market regimes. The project concludes with an interactive visualization framework for interpreting agent behavior and market dynamics.

1 Introduction

Financial markets are complex adaptive systems shaped by interactions among heterogeneous agents operating under uncertainty, latency, and information constraints. Traditional quantitative strategies often abstract away these mechanisms by operating directly on historical price series.

This project adopts a bottom-up approach by first constructing a realistic electronic exchange. Rather than assuming frictionless execution, the system explicitly models a limit order book, price–time priority matching, partial fills, and discrete event-driven time progression.

Multiple agent classes are introduced, including noise traders, momentum traders, market makers, and a reinforcement learning agent. The objective is not immediate profitability, but to examine whether learning agents can extract meaningful structure from microstructure-level interactions.

2 Methodology

The system is organized into four layers: market microstructure, agent design, reinforcement learning formulation, and evaluation.

2.1 Limit Order Book and Matching Engine

The simulator implements a centralized limit order book operating under price–time priority. Bid orders are stored in a max-heap, while ask orders are stored in a min-heap. Orders at identical price levels follow FIFO execution.

Market orders consume liquidity by walking the book until the requested quantity is filled or available liquidity is exhausted. Partial fills are supported, and all executions are recorded in a persistent trade log.

2.2 Event-Driven Market Simulation

Market evolution follows a discrete-event simulation framework. Time advances only when events occur, ensuring deterministic replay under fixed random seeds. Events include order arrivals, cancellations, snapshot recording, and market closure.

2.3 Agent Architecture

All agents inherit from a common abstract interface and interact with the market solely through action emission. Three non-learning agent types are implemented:

- **Noise Traders**: Zero-intelligence agents generating stochastic order flow.

- **Momentum Traders**: Trend-following agents reinforcing recent price movements.

- **Market Makers**: Liquidity providers posting bid–ask quotes with inventory-based skew.

This heterogeneous ecosystem enables realistic microstructure phenomena to emerge endogenously.

2.4 Reinforcement Learning Agent

The learning agent is formulated as an MDP and trained using PPO. The state vector includes market-level variables and portfolio information. The action space is discrete, and

the reward function reflects incremental PnL with risk penalties for inventory exposure and drawdowns.

3 Experimental Design and Results

3.1 Market Regime Experiments

Three controlled market regimes are studied by varying agent composition while holding all other parameters constant.

3.1.1 Noise Traders Only

The resulting price series resembles a random walk with wide and unstable bid–ask spreads, serving as a baseline regime.

3.1.2 Noise Traders with Market Makers

Liquidity provision stabilizes prices, tightens spreads, and induces mean-reverting behavior.

3.1.3 Noise Traders with Momentum Traders

Positive feedback leads to trend formation, volatility clustering, and abrupt reversals, illustrating endogenous instability.

3.2 Validation of Stylized Market Facts

3.2.1 Volatility Clustering

While raw returns exhibit weak autocorrelation, absolute returns display persistent positive autocorrelation, confirming volatility clustering.

3.2.2 Fat-Tailed Returns

Return distributions exhibit heavy tails relative to a Gaussian benchmark, indicating leptokurtic behavior consistent with empirical financial data.

3.3 Herding Dynamics

Position correlations among momentum traders increase sharply during stress periods, confirming that crashes are driven by synchronized behavior rather than isolated actions.

4 Empirical Results and Analysis

Results demonstrate that market structure alone can generate realistic price dynamics. Liquidity provision stabilizes markets, feedback traders amplify volatility, and RL agents learn risk-aware behavior under appropriate reward design.

5 Model Limitations and Scope

The simulator abstracts away several real-world complexities, including asymmetric information, hidden liquidity, and realistic latency modeling. Agent behavior is intentionally constrained to ensure interpretability and learning stability.

6 Conclusion

This project demonstrates that realistic market behavior and meaningful learning dynamics can emerge from disciplined system design. Rather than hard-coding outcomes, the simulator allows structure to arise endogenously from agent interaction and execution mechanics.

The reinforcement learning agent exhibits adaptive, risk-sensitive behavior without explicit strategy encoding, validating the MDP formulation. While results are not directly deployable, the framework provides a robust research instrument for studying market microstructure and learning-based trading systems.