Spectral Feature Extraction for Reinforcement Learning in Financial Markets Under Nonstationary Dynamics

Erik A. Bloomquist and Dr. Catia S. Silva
Department of Electrical & Computer Engineering, University of Florida

Introduction

Financial markets are **nonstationary**- their statistical properties change over time; learned strategies for autonomous trading must adapt across regimes as underlying distributions shift. **Reinforcement learning** (RL) offers a principled framework for sequential decision-making, but the relative impact of *reward engineering* vs *feature engineering* on policy performance remains underexplored. This study isolates both dimensions through a controlled factorial experiment.

If we are going to train an agent to trade stocks, will it perform better if we include **spectral** (frequency-domain) information? Are spectral features more useful in the **observation space** or in the **reward function**?

Methodology

Six RL agents are trained using the **Proximal Policy Optimization** (PPO) algorithm on daily SPY data from 2015-2024. Each agent observes one of three feature sets: raw momentum, spectral, or a combined dual-stream. Reward functions varied between a simple return-based baseline and a spectrally-informed reward that amplifies gains in trending markets and penalizes exposure during noisy, high-frequency periods.

Performance of each model is measured using several metrics, including total return, risk-adjusted metrics like Sharpe, Sortino, and Calmar (variants of Sharpe that penalize only losses and account for drawdown, respectively) Ratios, and maximum drawdown.

Model Condition	Reward Function	Feature Type	Observation Dim	Model Architecture
A	Simple Return	Raw Momentum	8	MLP 8-64-64
B	Simple Return	Spectral	4	MLP 4-64-64
C	Spectrally Informed	Raw Momentum	8	MLP 8-64-64
D	Spectrally Informed	Spectral	4	MLP 4-64-64
E	Simple Return	Dual-Stream	12	Dual-Encoder 2x (4-64), concat (128), MLP (64)
F	Spectrally Informed	Dual-Stream	12	Dual-Encoder 2x (4-64), concat (128), MLP (64)

Table 1. Summary of model conditions (A-F) under test.

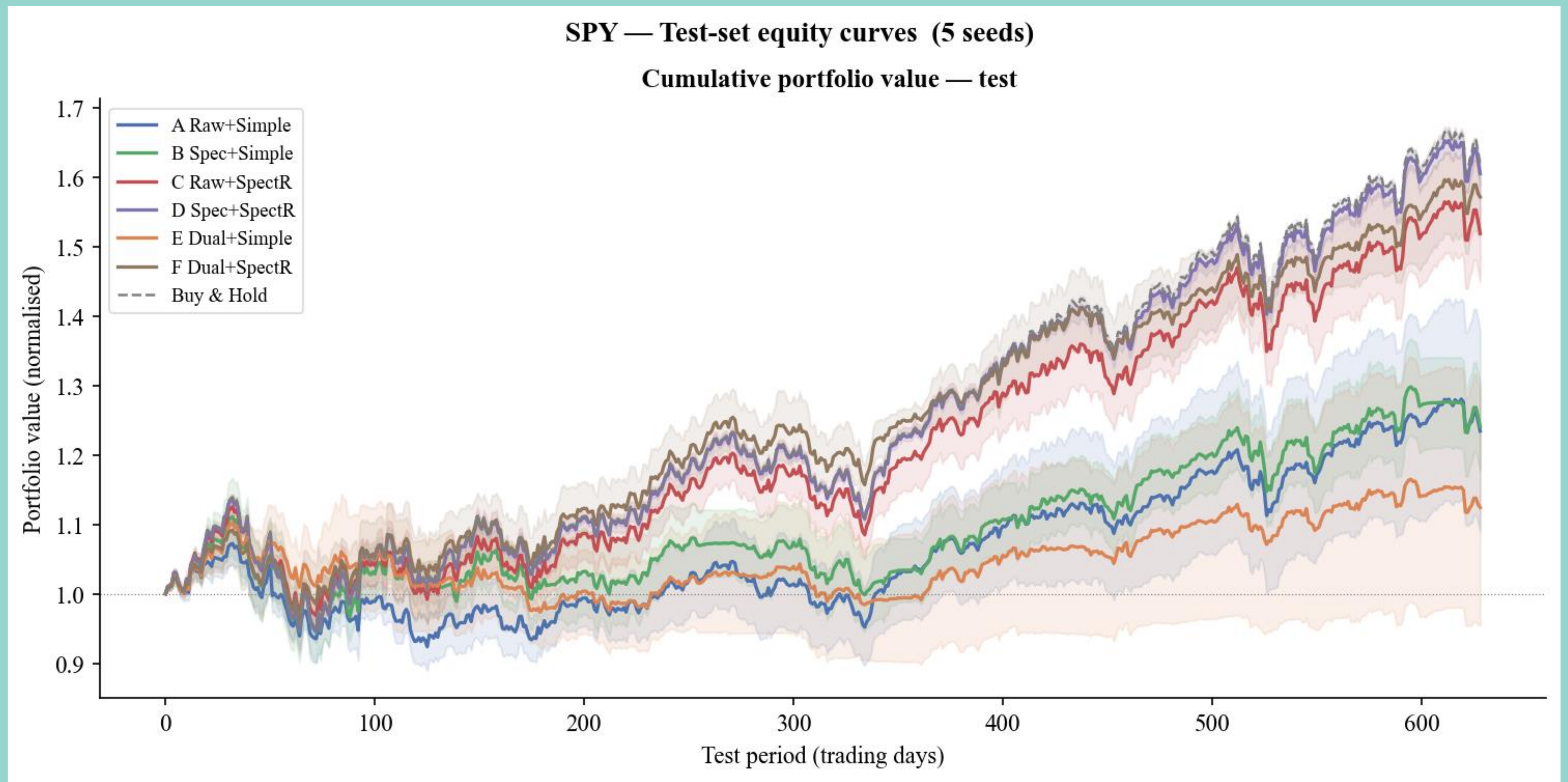

Figure 1. Test set equity curves for each model. The equity curve shows the value of an agent's portfolio relative to its starting value as days progress.

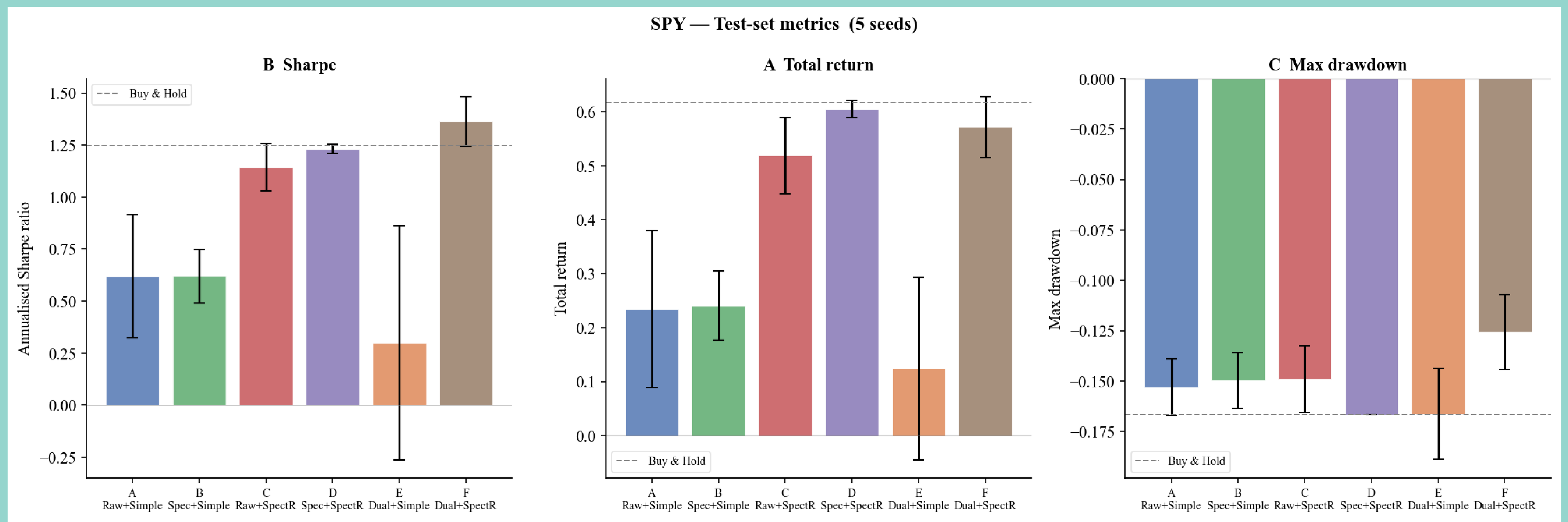

Figure 2. Various performance metrics for each model. Total return (A) measures the cumulative increase of an agent's portfolio over the test period. Sharpe Ratio (B) measures return relative to volatility- higher values indicate better risk-adjusted performance. Maximum drawdown (C) is the largest peak-to-trough decline in portfolio value over the evaluation period.

Results

Reward function was the dominant driver of performance. All three models using spectral reward outperformed their simple reward counterparts on every risk-adjusted metric (Sharpe, Sortino, Calmar Ratios).

Spectral features alone had no benefit. Conditions A and B (simple reward) achieved the same Sharpe Ratio (0.62), demonstrating that frequency-domain observations without an adapted reward signal provide no additional benefit.

Condition F (Dual-Stream features, Spectral reward) performed best. With a Sharpe Ratio of 1.36, Sortino Ratio of 1.93, and maximum drawdown of -12.6%, condition F outperformed the "buy & hold" baseline on all risk-adjusted metrics. Condition E (Dual-Stream features, Simple reward) showed severe overfitting, evidenced by its training Sharpe of 1.88 collapsing to a test Sharpe of 0.30.

Rank	Condition	Model Description	Sharpe (higher = better)	Max Drawdown (closer to 0 = better)
1	**F**	Dual-stream features, spectral reward	**1.36**	**-12.6%**
2	D	Spectral features, spectral reward	1.23	-16.7%
3	C	Momentum features, spectral reward	1.14	-14.9%
4	B	Spectral features, simple reward	0.62	-15.0%
5	A	Momentum features, simple reward	0.62	-15.3%
6	E	Dual-stream features, simple reward	0.30	-16.6%

Table 2. Test-period rankings by Sharpe Ratio. All three spectral reward conditions (F, D, C) outrank all simple reward conditions.

Conclusions and Future Directions

This study supports the following conclusions:

- Reward function design was the dominant determinant of test-period performance-- more impactful than choice of observation features
- All three spectral reward conditions (F, D, C) outperformed all three simple reward conditions on every risk-adjusted metric
- Spectral features added to the observation space alone provided no benefit-- Conditions A and B achieved identical Sharpe Ratios (0.62) despite differing feature sets
- Frequency-domain information is most valuable when embedded into the reward function, where it shapes credit assignment across market regimes

Future work should aim to:

- Extend the framework to multi-asset portfolio management
- Test whether spectral reward benefit persists in shorter-horizon or higher-volatility market conditions
- Investigate adaptive spectral reward weighting that updates hyperparameters in response to changing market conditions