UCLA Anderson School of Management

Advanced Data Analytics & AI

Deep Reinforcement Learning for Dynamic Stat Arb

Motivation & Problem

Why Classical Stat Arb Fails

- Correlations drift; cointegration breaks during regime shifts.
- Distance / OU-based spreads impose linearity and stationarity
- Fixed z-score thresholds cannot adapt to market volatility
- Signals are noisy when pair selection is based only on historical correlation

Our Objective

Replace the static stat-arb pipeline with a dynamic learning-based system that:

- Finds economically meaningful pairs
- Learns adaptive trading rules
- Reacts to regime shifts without fixed thresholds

Pipeline Overview

1. Pair Selection

- Build economic / sector relationship network
- Add supplier-customer, competitor edges
- Combine with Spearman correlations → top monthly in-sector pairs
- Output: a stable, economically grounded graph of valid pairs

2. Feature Engineering

- Hedge ratio β, spread, z-score, Δ-spread.
- Volatility (rolling std), half-life of mean reversion, residual spread (factor regression).

3. RL Trading Agent

- PPO - LSTM
- State = engineered features + current position
- Actions = Long / Flat / Short
- Reward = ΔPnL – transaction cost

4. Dynamic Backtest

- Each month: trade active pairs, aggregate PnL into a portfolio.

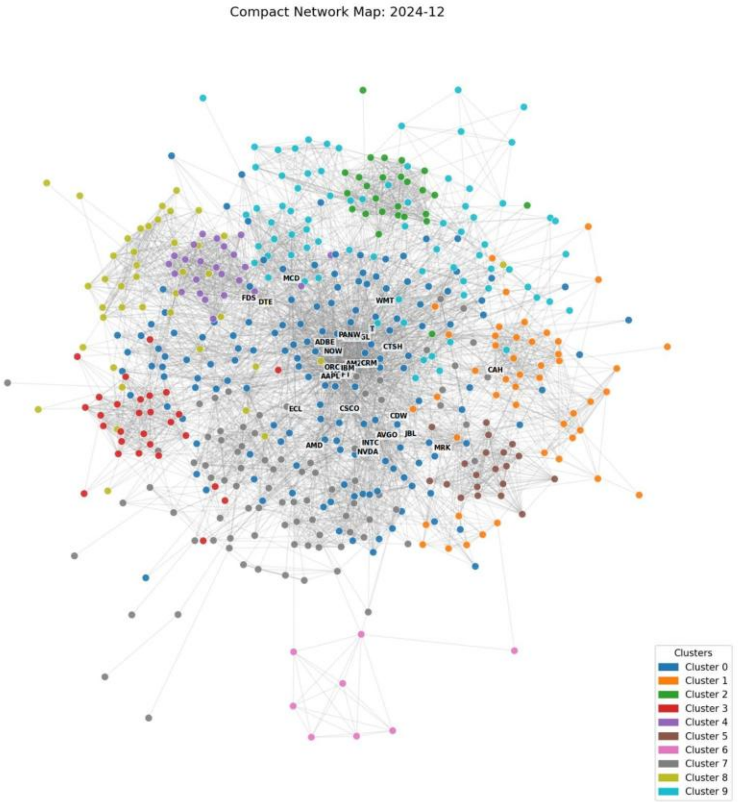
Pair Selection via Economic Network

Data sources: sector/subsector overlap, industry classifications and economic links.

- Build multi-layer graph:
- Industry & sub-industry co-membership
- Supplier customer & competitor edges (weighted)
- Convert into monthly adjacency matrix
- Compute Spearman correlations within each cluster
- Select the strongest non-duplicate pairs per cluster

Why this matters:

Produces stable, economically grounded pairs instead of noisy correlation screens



Features & RL Environment

Spread & Signal Engineering:

- Rolling hedge ratio β
- Log-spread, z-score, Δspread
- Volatility (20–30d), half-life
- Technical Factors (RSI, MACD, Bollinger)
- Last 4 days returns
- Residual spreads via ETF factor regression

→ Clean, de-factorized mean-reversion signals

RL Environment:

- **State**: all features + current position
- **Actions**: Short / Flat / Long
- **Reward**: ΔPnL
- **Agent**: Recurrent PPO (PPO-LSTM)
  (handles sequence structure)

Portfolio Backtest

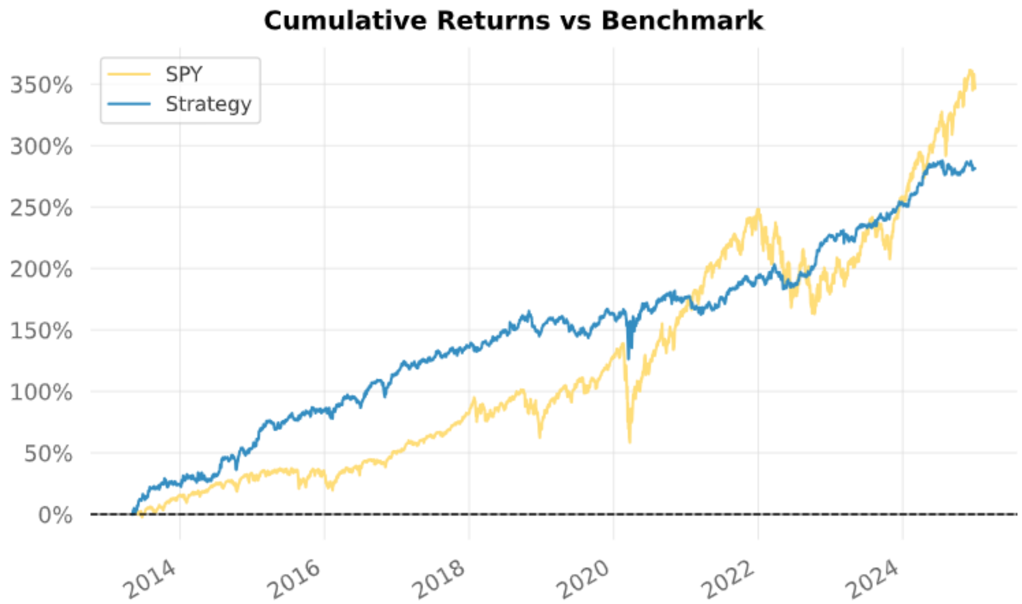
Dynamic trading setup:

- Monthly selection of top in-sector pairs.
- Allocate fixed capital per pairs
- Run daily inference with pair-specific PPO models
- Aggregate to portfolio equity

Key Performance Metrics

Metric	SPY	Strategy
Risk-Free Rate	0.0%	0.0%
Time in Market	100.0%	100.0%
Cumulative Return	346.51%	281.38%
CAGR%	13.71%	12.18%
Sharpe	0.85	1.27
Prob. Sharpe Ratio	99.77%	100.0%
Smart Sharpe	0.82	1.22
Sortino	1.19	1.96
Smart Sortino	1.14	1.89
Sortino/√2	0.84	1.39
Smart Sortino/√2	0.81	1.33
Omega	1.27	1.27

Portfolio Backtest



Limitations & Possible Extensions

1. Static Pair Universe (within month) → Rolling TIC Network

Refine pair selection using rolling persistence + graph stability.

2. Simple z-score spreads → EMRT & nonlinear β

Use model-free mean-reversion metrics (Ning & Lee, 2024)

3. Per-pair RL → Portfolio-level RL

Joint allocation decisions across spreads.

4. Simplified execution → Microstructure realism

Model slippage, spread widening, partial fills.

5. Single Agent PPO → Compare SAC/TD3

Evaluate stability & continuous sizing.

Thank You