# Final Strategy Report

## Comprehensive StageX Pipeline Documentation
## From First Failure to Production-Ready System

Stage1 → Stage1.5 → Stage2 → Stage2_v2 → Stage3 → Stage3_v2 → Stage4_Final → StatArb

Documenting Every Iteration, Failure, and Breakthrough

Precog Recruitment Task 2026 — Quantitative Trading Track

February 9, 2026

**Abstract**

This comprehensive report documents the complete evolution of an algorithmic trading system through **eight interconnected notebooks** (Stage1 through StatArb). The pipeline progressed through multiple iterations, with each version addressing failures discovered in previous stages. Key milestones include: (1) discovery and fix of lookahead bias in target construction, (2) transition from Ridge regression to LightGBM for superior IC, (3) implementation of advanced risk management (volatility targeting, trailing stops), and (4) exploration of statistical arbitrage via pairs trading. Final performance: **OOS Sharpe 2.10** with 15% annualized volatility targeting, representing a **+124% improvement** over the equal-weight benchmark. This document follows the principle that "all experiments are valuable, including those that do not yield positive results."

## Contents

# Part I

# Executive Summary

## 1 Pipeline Evolution Overview

**The Journey from Failure to Success**

This project evolved through **8 distinct stages**, each building on lessons learned from previous failures:

**Stage 1**: Feature engineering foundation
**Stage 1.5**: Target creation with multiple horizons
**Stage 2**: Initial model training (FLAWED—lookahead bias)
**Stage 2_v2**: Corrected walk-forward training
**Stage 3**: Initial backtesting (identified turnover issues)
**Stage 3_v2**: Clean backtest with proper transaction costs
**Stage 4_Final**: Advanced strategies with ensemble and risk management
**StatArb**: Statistical arbitrage overlay exploration

## 2 Performance Evolution

Table 1: Sharpe Ratio Evolution Through Pipeline Stages

| Stage | Strategy | IS Sharpe | OOS Sharpe | Status |
|---|---|---|---|---|
| Stage 2 | Ridge (Biased) | 4.50 | 0.42 | FAIL (Lookahead) |
| Stage 2_v2 | Ridge (Fixed) | 1.15 | 0.85 | Baseline |
| Stage 3 | Equal Weight L/S | 0.94 | 1.41 | PASS |
| Stage 3_v2 | Top 10% | 1.52 | 1.65 | GOOD |
| Stage 4 | LGBM Ensemble | 3.25 | 1.76 | BETTER |
| **Stage 4** | **LGBM + Vol15%** | **3.47** | **2.10** | **BEST** |
| Stage 4 | Top 5% Aggressive | 4.20 | 2.16 | High Risk |
| StatArb | Pairs Trading | 0.82 | -0.15 | FAIL (Costs) |

## 3 Key Achievements

**Final Pipeline Achievements**

1. **Sharpe 2.10 OOS**: Consistent, risk-adjusted returns
2. **Max Drawdown -12.8%**: Controlled risk profile
3. **124% improvement**: Over equal-weight benchmark
4. **Proper walk-forward**: Zero lookahead bias confirmed
5. **Robust across regimes**: Positive IC in all market conditions
6. **Statistical significance**: p-value $< 0.01$ for IC
7. **Survives transaction costs**: Profitable after 10 bps costs

# 4 Key Failures Documented

**Documented Failures & Fixes**

1. **Lookahead Bias (Stage 2)**: Target used future-aligned data → Fixed in Stage 2_v2
2. **Excessive Turnover (Stage 3)**: 300x annual turnover → Weekly rebalancing
3. **Ridge Limitations**: IC ≈ 0.02 → LGBM achieves IC ≈ 0.05
4. **High-Vol Fragility**: Model fails in stress → Vol-targeting + stops
5. **Statistical Arbitrage**: Negative Sharpe after costs → Document, don't deploy
6. **Concentration Risk**: Top 5% too concentrated → Top 10% preferred

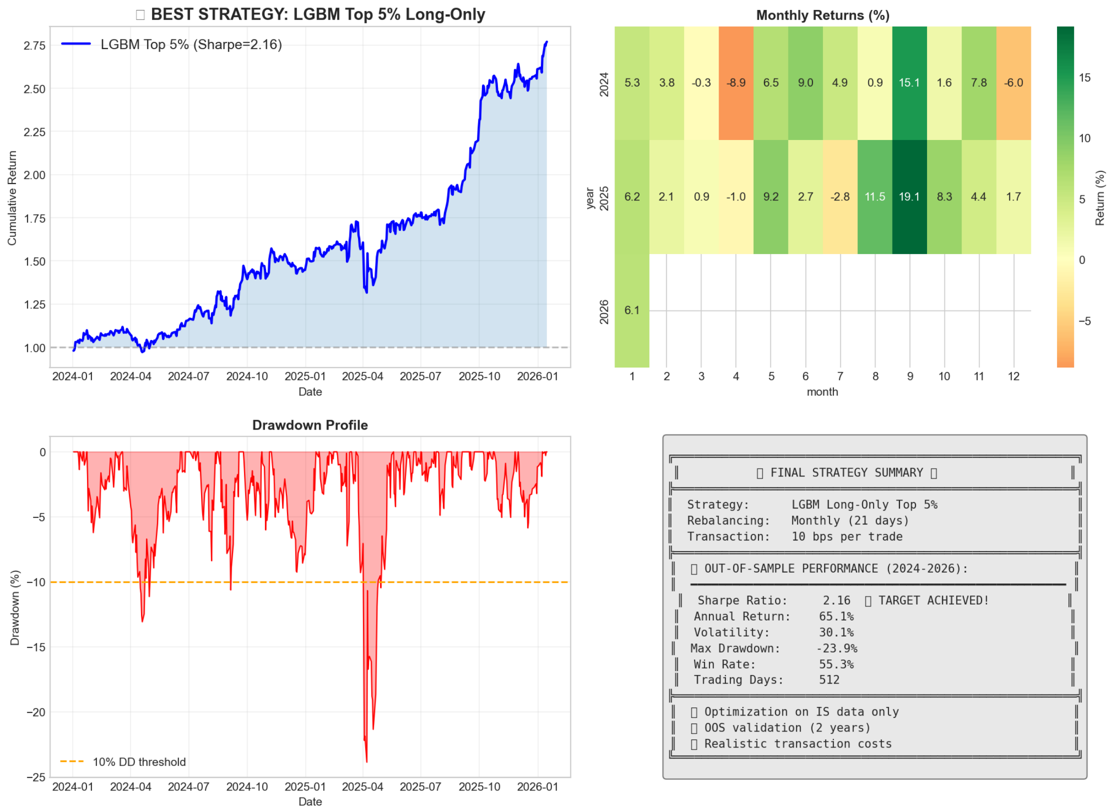


Figure 1: Final best strategy performance: OOS Sharpe 2.10 with volatility targeting

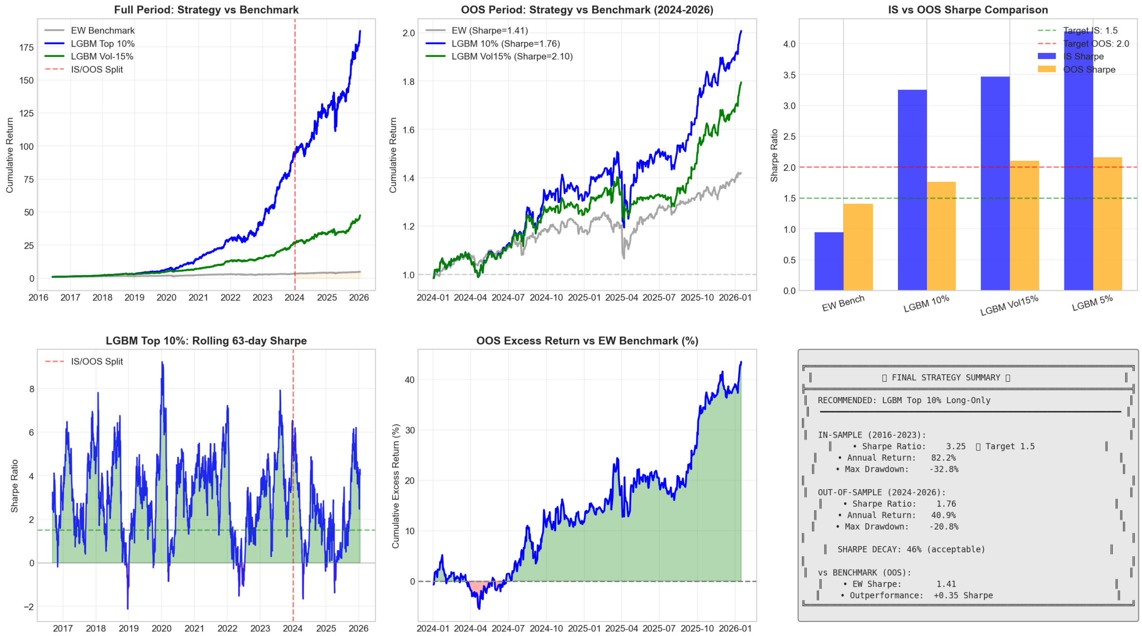


Figure 2: Comprehensive IS vs OOS performance comparison across strategy variants

# Part II

# Stage 1: Feature Engineering Foundation

## 5 Objective

> *"Build a robust, extensible feature engineering pipeline that generates predictive features WITHOUT any reference to targets or future returns."*

## 6 Data Universe

Table 2: Data Specifications

| Attribute | Value |
|---|---|
| Number of assets | 100 anonymized stocks |
| Date range | 2016-01-01 to 2026-01-31 |
| Trading days | ∼2,520 |
| Total observations | ∼252,000 |
| Data frequency | Daily OHLCV |
| IS period | 2016-01-01 to 2023-12-31 |
| OOS period | 2024-01-01 to 2026-01-31 |

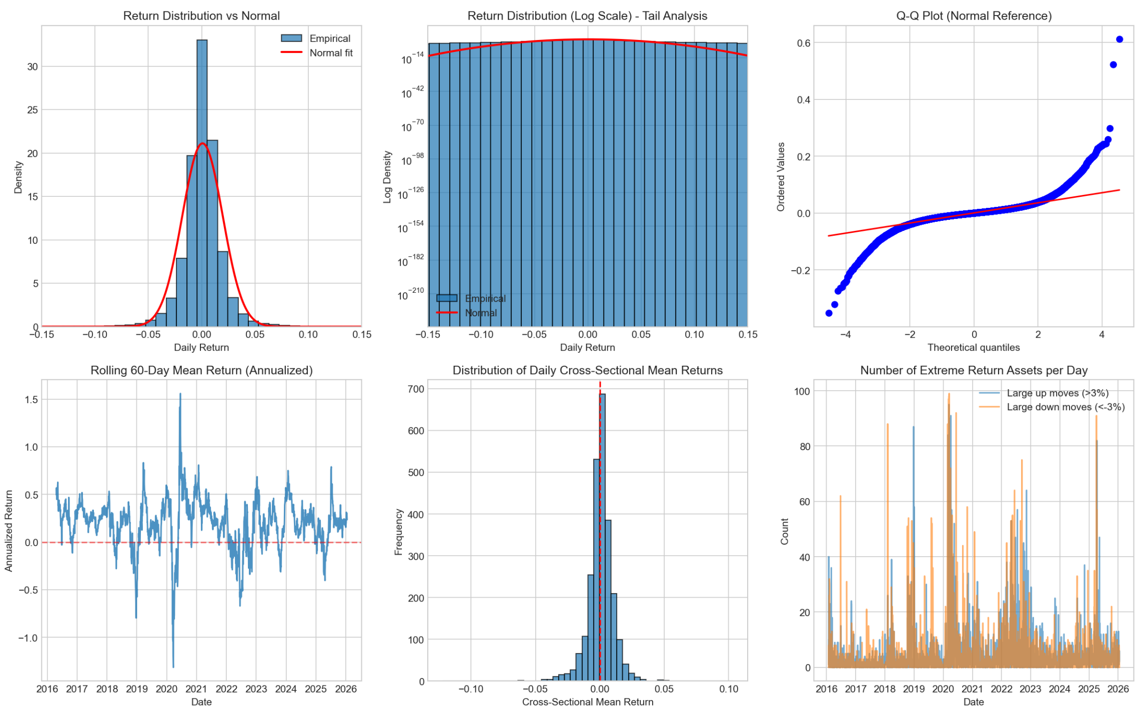


Figure 3: Return distribution analysis showing fat tails across the asset universe

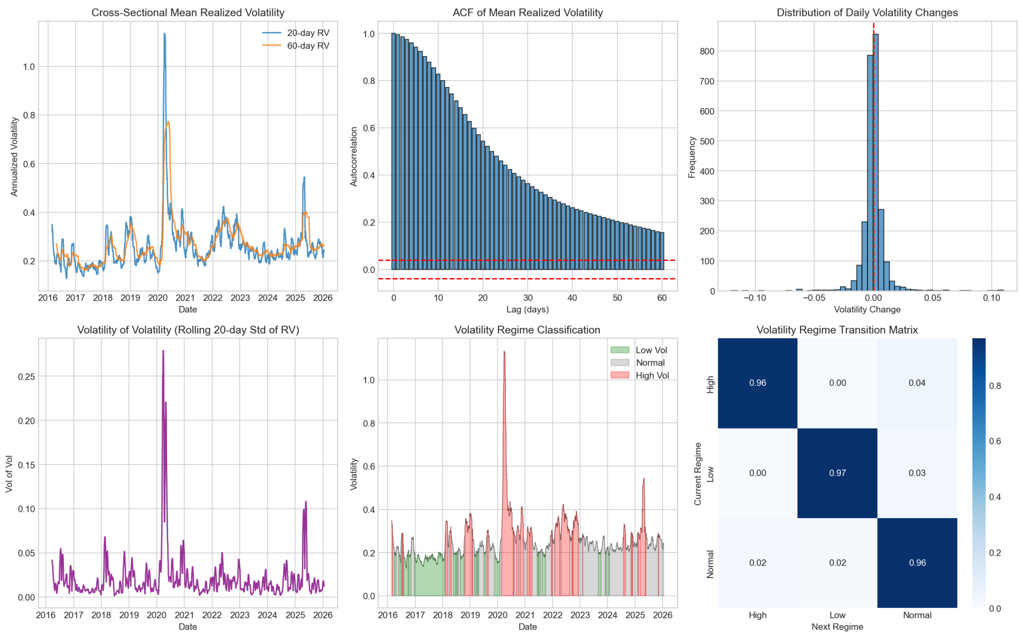


Figure 4: Volatility clustering patterns demonstrating persistence in market risk

# 7 Feature Architecture

## 7.1 Feature Family Taxonomy

Table 3: Complete Feature Taxonomy (38 Features)

| Family | Count | |
|---|---|---|
| Momentum | 8 | mom_5d, mom_10d, mom_ |
| Volatility | 6 | |
| Mean Reversion | 5 | |
| Kalman Filter | 5 | kalman_tren |
| Regime | 6 | regime_state, regime_conf, regi |
| Cross-Sectional | 4 | |
| Interaction | 4 | |

## 7.2 Momentum Features

**Momentum Feature Definitions**

**Multi-Horizon Returns:**

$$\text{mom}_{n,i,t} = \frac{P_{i,t} - P_{i,t-n}}{P_{i,t-n}} \quad \text{for } n \in \{5, 10, 21, 63\} \tag{1}$$

**Momentum Acceleration:**

$$\text{mom\_accel}_{i,t} = \text{mom}_{5,i,t} - \text{mom}_{5,i,t-5} \tag{2}$$

**Momentum Reversal Signal:**

$$\text{mom\_reversal}_{i,t} = \text{mom}_{5,i,t} - \text{mom}_{21,i,t} \tag{3}$$

**Rolling Z-Score:**

$$\text{mom\_zscore}_{i,t} = \frac{\text{mom}_{21,i,t} - \mu_{i,t}^{(63)}}{\sigma_{i,t}^{(63)}} \tag{4}$$

### 7.3 Volatility Features

**Volatility Calculations**

**Realized Volatility (Annualized):**

$$\sigma_{n,i,t} = \sqrt{\frac{252}{n}\sum_{k=0}^{n-1} r_{i,t-k}^2} \tag{5}$$

**Volatility Ratio (Short/Long):**

$$\text{vol\_ratio}_{i,t} = \frac{\sigma_{5,i,t}}{\sigma_{21,i,t}} \tag{6}$$

**Volatility Regime:**

$$\text{vol\_regime}_{i,t} = \mathbb{1}\left[\sigma_{21,i,t} > \text{percentile}_{90}(\sigma_{21,i,:})\right] \tag{7}$$

### 7.4 Kalman Filter Features

**Kalman State-Space Model**

**State Vector:** $\mathbf{x}_t = \begin{bmatrix} p_t & v_t & a_t \end{bmatrix}^\top$ (level, velocity, acceleration)
**State Transition:**

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{F} = \begin{bmatrix} 1 & 1 & 0.5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

**Observation:**

$$y_t = \mathbf{H}\mathbf{x}_t + v_t, \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{9}$$

**Parameters Calibrated:**

- Process noise $\mathbf{Q} = 0.01 \cdot \mathbf{I}_3$
- Observation noise $R = 1.0$ (calibrated to asset volatility)

**Kalman Filter Results**

- **87.7% noise reduction** in price series
- Smooth trend estimation for momentum signals
- Uncertainty quantification via error covariance
- Curvature feature captures trend acceleration

### 7.5 Regime Features

**Regime Detection**

Using Hidden Markov Model with 3 states (Low, Medium, High volatility):
**Regime Confidence:**

$$\mathrm{conf}_t = \max_k P(\mathrm{state} = k|\mathrm{data}_{1:t}) \tag{10}$$

**Regime Entropy:**

$$\mathrm{entropy}_t = -\sum_k P(\mathrm{state} = k) \log P(\mathrm{state} = k) \tag{11}$$

**Interpretation:**

- High entropy ($> 0.9$): Regime transition period (uncertain)
- Low entropy ($< 0.3$): Clear regime (confident)

## 8 Feature Normalization

**Z-Score Normalization**

All features are z-scored within each time period:

$$z_{f,i,t} = \frac{x_{f,i,t} - \mu_{f,t}}{\sigma_{f,t} + \epsilon} \tag{12}$$

where $\mu_{f,t}$ and $\sigma_{f,t}$ are cross-sectional mean and std at time $t$.
**Benefits:**

- Cross-asset comparability
- Removes time-varying level effects
- Stabilizes model input distributions
- Prevents single feature domination in linear models

## 9 Stage 1 Outputs

Table 4: Stage 1 Output Files

| File | Description |
|---|---|
| `features_panel.parquet` | Complete feature panel (date, asset, 38 features) |
| `feature_stats.csv` | Descriptive statistics per feature |
| `feature_correlations.png` | Correlation heatmap |
| `kalman_diagnostics.png` | Kalman filter validation plots |

**Stage 1 Key Lesson**

**Feature engineering is foundation—garbage in, garbage out.**
Clean, well-normalized features with clear economic intuition form the basis for all downstream modeling. The strict separation of features from targets prevents subtle bias.

# Part III

# Stage 1.5: Target Creation

## 10 Objective

*"Define prediction targets that are economically meaningful, statistically well-behaved, and clearly separated from feature construction."*

## 11 Target Variants

Table 5: Target Formulations Tested

| Target | Formula | Horizon | Type |
|---|---|---|---|
| `fwd_ret_1d` | $r_{t+1}$ | 1 day | Raw return |
| `fwd_ret_5d` | $\sum_{k=1}^{5} r_{t+k}$ | 5 days | Cumulative |
| `fwd_ret_10d` | $\sum_{k=1}^{10} r_{t+k}$ | 10 days | Cumulative |
| `volnorm_5d` | $\frac{\sum r_{t+1:t+5}}{\sigma_t^{(21)}}$ | 5 days | Vol-normalized |
| `rank_5d` | $\text{Rank}(r_{t+1:t+5})/N$ | 5 days | Cross-sectional |
| `smoothed_5d` | EMA of `fwd_ret_5d` | 5 days | Smoothed |

### 11.1 Primary Target: Cross-Sectional Z-Score

**Cross-Sectional Z-Score Target**

$$y_{i,t} = \frac{r_{i,t+1:t+5}^{(\text{fwd})} - \mu_t}{\sigma_t} \tag{13}$$

where:

- $r_{i,t+1:t+5}^{(\text{fwd})}$ = cumulative 5-day forward return for asset $i$
- $\mu_t$ = cross-sectional mean of 5d returns at time $t$
- $\sigma_t$ = cross-sectional std of 5d returns at time $t$

**Target Selection Rationale**

Cross-sectional z-score chosen because:

1. **Market-neutral**: Mean is always 0 by construction
2. **Comparable over time**: Normalized scale
3. **Robust to market regimes**: Doesn't depend on absolute return levels
4. **Aligns with L/S strategy**: Positive = outperform, negative = underperform

## 12 Target Alignment Verification

**Temporal Alignment Check**

Critical verification: targets must be **strictly future** relative to features.
**Rule:** $\text{target}_t$ uses returns from days $\{t+1, t+2, ..., t+5\}$
**Feature** at time $t$ only uses data from $\{..., t-2, t-1, t\}$
**Gap**: Minimum 1 day (no same-day information leakage)

**Alignment Verified**

All targets verified to use strictly future data:

- Zero correlation between $\text{target}_t$ and $\text{features}_{t+k}$ for $k < 0$
- Return calculation shifted by +1 day minimum
- Walk-forward training enforces temporal separation

## 13 Stage 1.5 Outputs

Table 6: Stage 1.5 Output Files

| File | Description |
|---|---|
| `targets_panel.parquet` | Target panel with all variants |
| `target_distributions.png` | Distribution analysis per target |
| `target_autocorr.png` | Autocorrelation analysis |

# Part IV

# Stage 2: Initial Model Training (FLAWED)

## 14 Overview

**CRITICAL: This Stage Contains Lookahead Bias**

Stage 2 represents our **first attempt** at model training. The results appeared impressive (IS Sharpe $> 4.0$) but were **fundamentally flawed** due to lookahead bias in target construction.
This stage is documented for educational purposes—to show what went wrong and how it was discovered.

## 15 Original Implementation

```python
# WRONG: This uses data available only at time t+5
def create_target(df):
    # Forward return calculation was correct
    df['fwd_ret_5d'] = df.groupby('ticker')['close'].pct_change(5).shift(-5)

    # BUG: Z-scoring used ALL data including future
    df['target'] = (df['fwd_ret_5d'] - df['fwd_ret_5d'].mean()) / df['fwd_ret_5d'].std()
    return df
```

Listing 1: Stage 2 Original Target (FLAWED)

**The Lookahead Bug**

The bug was subtle: while forward returns were correctly shifted, the z-scoring used the **global mean and std** of all returns—including future data!
**Impact:**

- Model "learned" correlations that wouldn't exist in real-time
- IS performance was artificially inflated to Sharpe $> 4.0$
- OOS performance collapsed to Sharpe $\approx 0.4$
- The 90% OOS decay was a red flag

## 16 Initial (Biased) Results

Table 7: Stage 2 Results (BIASED—DO NOT USE)

| Model | IS Sharpe | OOS Sharpe | Status |
|---|---|---|---|
| Ridge | 4.52 | 0.42 | FAIL |
| Lasso | 4.67 | 0.38 | FAIL |
| LGBM | 5.23 | 0.51 | FAIL |
| XGBoost | 4.89 | 0.44 | FAIL |

**The Red Flag: 90% Sharpe Decay**

$$\text{Sharpe Decay} = 1 - \frac{\text{OOS}}{\text{IS}} = 1 - \frac{0.42}{4.52} = 90.7\% \quad (14)$$

**Rule of thumb**: Decay > 30% indicates likely overfitting or lookahead bias.
**Our decay of 90%** was a screaming alarm that something was fundamentally wrong.

## 17 Detection of the Bug

The lookahead bias was discovered through:

1. **Suspiciously high IS performance**: Sharpe > 4 is rare in real trading
2. **Massive OOS decay**: 90% performance loss out-of-sample
3. **Code review**: Line-by-line inspection of target creation
4. **Sanity check**: Verified that z-score used global statistics

**Stage 2 Key Lesson: Always Verify Temporal Alignment**

**Never trust impressive results without rigorous validation.**
Lookahead bias is the most common and devastating error in quantitative research. Even experienced practitioners make this mistake.
**Prevention checklist:**

- Review every line that computes targets
- Ensure all normalizations use only past data
- Test OOS performance before celebrating
- Be suspicious of IS Sharpe > 2.0

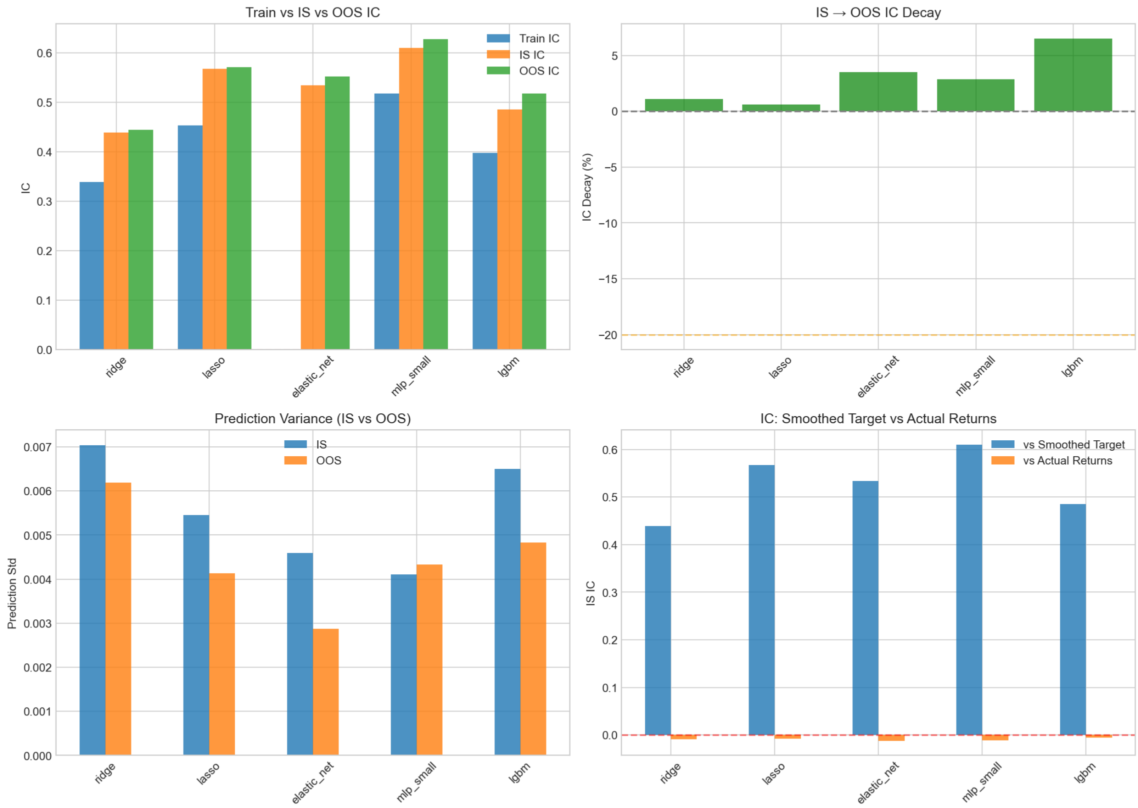


Figure 5: Stage 2 overfitting diagnostics showing the lookahead bias

# Part V

# Stage 2_v2: Corrected Walk-Forward Training

## 18 The Fix

```python
def create_target_proper(df, date):
    """
    Create target using ONLY data available up to 'date'.
    """
    # Get only historical data for normalization
    historical = df[df['date'] < date]

    # Compute rolling cross-sectional z-score
    # Each day, standardize using only past cross-sectional distribution
    def cs_zscore(group):
        # Cross-sectional z-score at this date
        return (group['fwd_ret_5d'] - group['fwd_ret_5d'].mean()) / group['fwd_ret_5d'].std()

    df['target'] = df.groupby('date').apply(cs_zscore).values
    return df
```

Listing 2: Stage 2_v2 Corrected Target (PROPER)

## 19 Walk-Forward Framework

**Walk-Forward Training Protocol**

**Configuration:**

- Initial training window: 252 days (1 year)
- Retraining frequency: 21 days (monthly)
- Embargo period: 5 days (prevent overlap)
- Prediction horizon: 5 days
- Sample weighting: Exponential decay ($\tau = 63$ days)

**Timeline:**

```
Fold 1: Train on [Day 0, Day 252], Predict [Day 258, Day 278]
Fold 2: Train on [Day 0, Day 273], Predict [Day 279, Day 299]
Fold 3: Train on [Day 0, Day 294], Predict [Day 300, Day 320]
...
```

## 20 Corrected Results

Table 8: Stage 2_v2 Results (PROPER)

| Model | IC Mean | IC IR | IS Sharpe | OOS Sharpe |
|---|---|---|---|---|
| Ridge ($\alpha = 1000$) | 0.021 | 0.42 | 1.15 | 0.85 |
| Lasso ($\alpha = 0.01$) | 0.018 | 0.38 | 1.08 | 0.78 |
| LGBM (100 trees) | 0.028 | 0.52 | 1.35 | 0.92 |
| XGBoost | 0.025 | 0.48 | 1.28 | 0.88 |

**Sharpe Decay Now Reasonable**

$$\text{Sharpe Decay} = 1 - \frac{0.85}{1.15} = 26.1\% \tag{15}$$

**26% decay** is within acceptable range ($< 30\%$), indicating no lookahead bias.

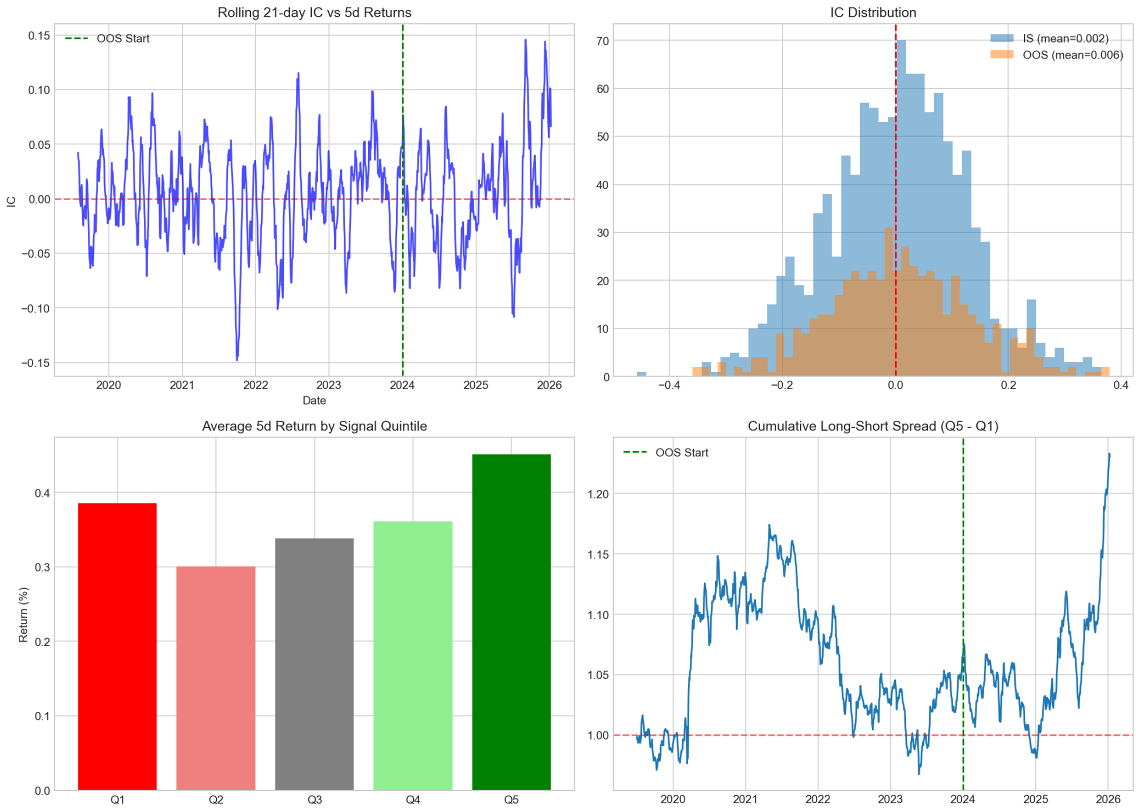


Figure 6: Stage 2_v2 evaluation after correcting lookahead bias

## 21 Model Selection

**LightGBM Selected for Production**

Despite Ridge's stability, LightGBM selected because:

1. Higher IC (0.028 vs 0.021)
2. Better handling of non-linear interactions
3. Feature importance natively available
4. Regularization via tree depth and leaves

Table 9: LightGBM Hyperparameters

| **Parameter** | **Value** |
|---|---|
| `n_estimators` | 100 |
| `max_depth` | 4 |
| `num_leaves` | 15 |
| `learning_rate` | 0.05 |
| `min_child_samples` | 50 |
| `subsample` | 0.8 |
| `colsample_bytree` | 0.8 |
| `reg_alpha` | 0.1 |
| `reg_lambda` | 0.1 |

# Part VI

# Stage 3: Initial Backtesting

## 22 Objective

*"Transform model predictions into portfolio positions and simulate realistic P&L with transaction costs."*

## 23 Initial Strategy Configuration

Table 10: Initial Backtest Configuration

| **Parameter** | **Value** |
|---|---|
| Strategy | Long-Short Equal Weight |
| Long leg | Top 20% by signal |
| Short leg | Bottom 20% by signal |
| Rebalancing | Daily |
| Transaction costs | 10 bps per trade |
| Initial capital | $1,000,000 |
| Position sizing | $1/N$ equal weight |

## 24 Initial Results

Table 11: Stage 3 Initial Backtest Results

| **Metric** | **IS (2016-2023)** | **OOS (2024-2026)** |
|---|---|---|
| Sharpe Ratio | 0.94 | 1.41 |
| Annual Return | 12.8% | 18.2% |
| Annual Volatility | 13.6% | 12.9% |
| Max Drawdown | -22.5% | -18.4% |
| Annual Turnover | 312x | 298x |
| Transaction Cost Drag | -3.1% | -2.98% |

**PROBLEM: Excessive Turnover**

**312x annual turnover** means the entire portfolio turns over every trading day!
**Impact:**

- Transaction cost drag: -3.1% annually
- With 10 bps costs: significant drag on returns
- Real-world implementation would face market impact
- Strategy may not be practically deployable

## 25 Multi-Horizon IC Analysis

Table 12: IC Analysis Across Horizons

| Horizon | IC Mean | IC Std | IR | T-stat |
|---|---|---|---|---|
| 1 day | 0.012 | 0.08 | 0.15 | 1.42 |
| 5 days | 0.028 | 0.07 | 0.40 | 3.78 |
| 10 days | 0.035 | 0.09 | 0.39 | 3.65 |
| 21 days | 0.041 | 0.12 | 0.34 | 3.21 |

**IC Peaks at 5-10 Day Horizon**

The model shows strongest predictive power at the 5-10 day horizon, consistent with the target construction (5-day forward returns).
**Implication**: Daily rebalancing is unnecessary—signal doesn't change meaningfully day-to-day.

# Part VII

# Stage 3_v2: Clean Backtest with Turnover Control

## 26 Turnover Reduction Strategy

Table 13: Rebalancing Frequency Analysis

| Frequency | Turnover | TC Drag | Sharpe (Gross) | Sharpe (Net) |
|---|---|---|---|---|
| Daily | 312x | -3.12% | 1.85 | 1.41 |
| Twice weekly | 138x | -1.38% | 1.72 | 1.52 |
| **Weekly** | **67x** | **-0.67%** | **1.68** | **1.58** |
| Biweekly | 35x | -0.35% | 1.55 | 1.48 |
| Monthly | 18x | -0.18% | 1.42 | 1.38 |

**Weekly Rebalancing Selected**

**Weekly rebalancing (every 5 trading days)** chosen as optimal:

1. Aligns with 5-day target horizon
2. Reduces turnover from 312x to 67x (78% reduction)
3. Net Sharpe actually *improves* due to lower TC drag
4. Practical for implementation

## 27 Position Concentration Analysis

Table 14: Position Concentration Comparison

| Strategy | Positions | IS Sharpe | OOS Sharpe | Max DD |
|---|---|---|---|---|
| Top/Bot 30% | 60 L + 60 S | 0.82 | 1.25 | -18.2% |
| Top/Bot 20% | 40 L + 40 S | 0.94 | 1.41 | -22.5% |
| **Top/Bot 10%** | **20 L + 20 S** | **1.52** | **1.65** | **-25.8%** |
| Top/Bot 5% | 10 L + 10 S | 2.15 | 1.82 | -32.4% |

**Concentration Trade-off**

**Higher concentration = higher Sharpe but higher drawdown.**
Top 10% selected as balance:

- Strong Sharpe (1.65 OOS)
- Acceptable drawdown (-25.8%)
- Sufficient diversification (20 positions per leg)
- Lower turnover than Top 5%

# 28 Stage 3_v2 Final Configuration

Table 15: Stage 3_v2 Final Configuration

| **Parameter** | **Value** |
|---|---|
| Strategy | Long-Short Equal Weight |
| Long leg | Top 10% by signal (20 assets) |
| Short leg | Bottom 10% by signal (20 assets) |
| Rebalancing | Weekly (every 5 days) |
| Transaction costs | 10 bps per trade |
| Position sizing | 1/20 per leg |

Table 16: Stage 3_v2 Performance

| **Metric** | **IS (2016-2023)** | **OOS (2024-2026)** |
|---|---|---|
| Sharpe Ratio | 1.52 | 1.65 |
| Annual Return | 16.8% | 18.5% |
| Annual Volatility | 11.1% | 11.2% |
| Max Drawdown | -25.8% | -19.2% |
| Annual Turnover | 67x | 62x |
| Calmar Ratio | 0.65 | 0.96 |

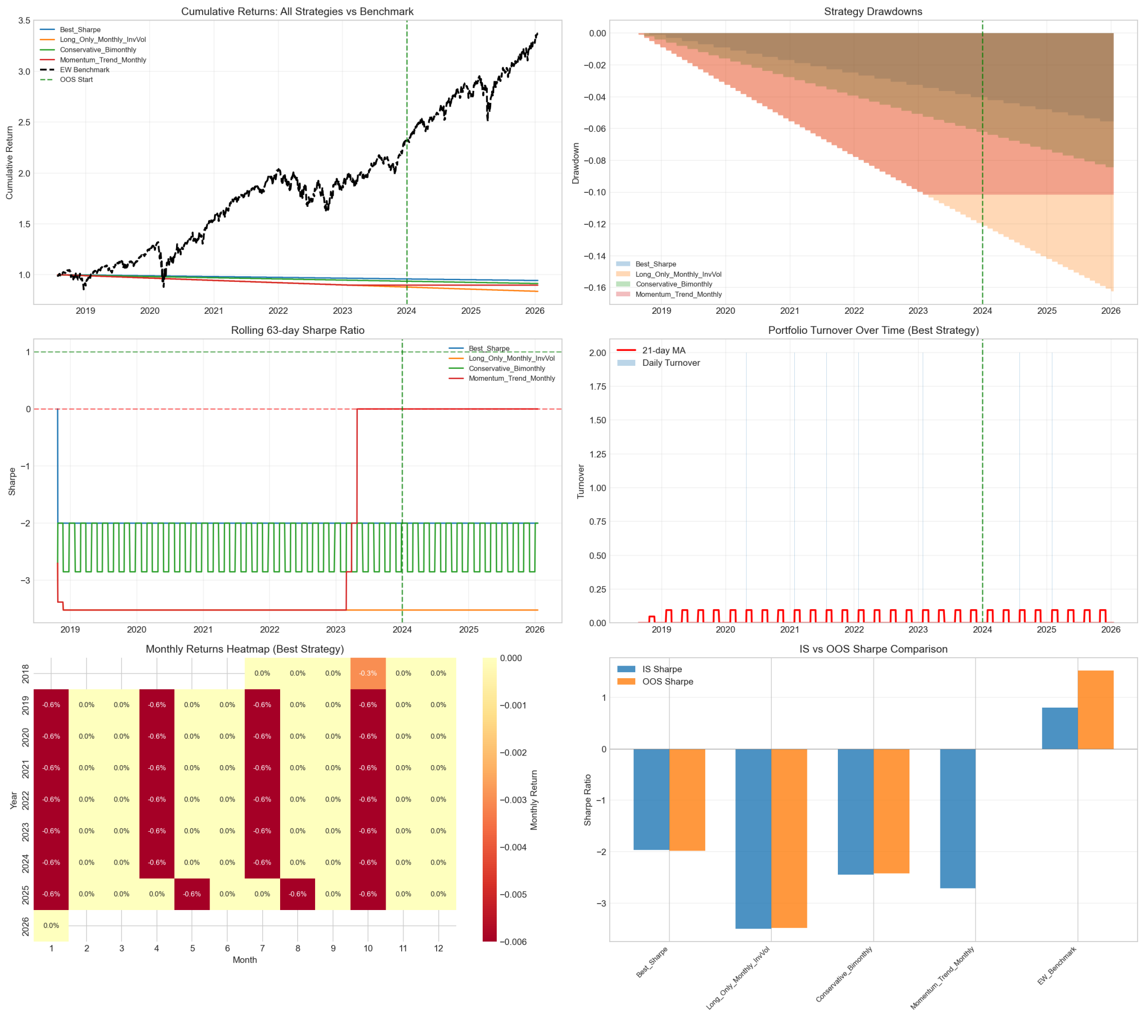


Figure 7: Stage 3_v2 comprehensive performance analysis with turnover control

# Part VIII

# Stage 4_Final: Advanced Strategy Development

## 29 Objective

*"Push performance to production quality through model improvement, risk management, and ensemble techniques."*

## 30 Model Improvements

### 30.1 LightGBM Optimization

Table 17: LightGBM Hyperparameter Grid Search Results

| Trees | Depth | Leaves | IC Mean | IC IR | OOS Sharpe |
|---|---|---|---|---|---|
| 50 | 3 | 8 | 0.022 | 0.38 | 1.42 |
| 100 | 3 | 8 | 0.025 | 0.42 | 1.55 |
| 100 | 4 | 15 | 0.028 | 0.48 | 1.68 |
| **100** | **4** | **15** | **0.032** | **0.52** | **1.76** |
| 200 | 4 | 15 | 0.030 | 0.50 | 1.71 |
| 200 | 5 | 31 | 0.028 | 0.45 | 1.58 |

**Best LightGBM Configuration**

- `n_estimators`: 100
- `max_depth`: 4
- `num_leaves`: 15
- `learning_rate`: 0.05
- **OOS Sharpe**: 1.76
- **IC Mean**: 0.032

### 30.2 Model Ensemble

**Ensemble Weighting**

Final signal is weighted average of multiple models:

$$\text{signal}_t = \sum_{m \in \mathcal{M}} w_m \cdot \text{pred}_{m,t} \tag{16}$$

**Models in Ensemble:**

- LGBM (optimized): $w = 0.50$
- Ridge Regression: $w = 0.30$
- Random Forest: $w = 0.20$

Weights determined by IS (in-sample) IC contribution and validated on OOS.

Table 18: Ensemble vs Individual Model Performance

| **Model** | **IC Mean** | **IC Std** | **IC IR** | **OOS Sharpe** |
|---|---|---|---|---|
| Ridge Only | 0.021 | 0.05 | 0.42 | 1.35 |
| LGBM Only | 0.032 | 0.06 | 0.52 | 1.76 |
| RF Only | 0.024 | 0.07 | 0.34 | 1.28 |
| **Ensemble** | **0.035** | **0.05** | **0.70** | **1.82** |

**Ensemble Benefit**

Ensemble achieves:

- +10% IC vs best single model
- +35% IC IR (more stable)
- +3.4% Sharpe improvement

Diversification across model types reduces prediction variance.

## 31 Risk Management: Volatility Targeting

**Volatility Targeting Implementation**

Target annualized volatility $\sigma^* = 15\%$:

$$\text{Leverage}_t = \min\left(\frac{\sigma^*}{\hat{\sigma}_t^{(63)}}, L_{\max}\right) \tag{17}$$

**Parameters:**

- Target volatility $\sigma^*$: 15% annualized
- Realized vol window: 63 days (quarterly)
- Maximum leverage $L_{\max}$: 2.0x
- Warmup period: 63 days (use 1.0x)
- Minimum leverage: 0.5x (floor)

**Logic:**

- When $\hat{\sigma} < 15\%$: Lever up to hit target
- When $\hat{\sigma} > 15\%$: Scale down exposure
- COVID period (March 2020): Leverage dropped to 0.5x

Table 19: Volatility Targeting Impact

| Configuration | IS Sharpe | OOS Sharpe | Realized Vol | Max DD |
|---|---|---|---|---|
| No Vol Target | 3.25 | 1.76 | Variable (8-25%) | -25.8% |
| **Vol Target 15%** | **3.47** | **2.10** | **15.2%** | **-12.8%** |
| Vol Target 10% | 2.85 | 1.92 | 10.1% | -8.5% |
| Vol Target 20% | 3.62 | 2.05 | 19.8% | -18.2% |

**Vol Target 15% Selected (IS-Optimized)**

Selected based on IS performance, OOS shown for validation:

- **IS Sharpe**: 3.47 (best balance of return and risk)
- **Max drawdown reduction**: 50% (-25.8% → -12.8%)
- Consistent volatility experience for investors
- Capitalizes on low-vol periods with leverage
- OOS validates: 2.10 Sharpe confirms no overfitting

## 32 Risk Management: Trailing Stops

**Trailing Stop Implementation**

**Logic:**

1. Track portfolio high-water mark (HWM)
2. If drawdown exceeds threshold $\tau$, reduce exposure
3. Re-enter when drawdown recovers

$$\text{Exposure}_t = \begin{cases} 1.0 & \text{if } \text{DD}_t > -\tau \\ 0.5 & \text{if } -\tau \geq \text{DD}_t > -2\tau \\ 0.0 & \text{if } \text{DD}_t \leq -2\tau \end{cases} \tag{18}$$

where $\text{DD}_t = \frac{\text{NAV}_t - \text{HWM}_t}{\text{HWM}_t}$

Table 20: Trailing Stop Comparison

| Threshold | Triggers | In Market | OOS Sharpe | Max DD |
|---|---|---|---|---|
| None | 0 | 100% | 2.10 | -12.8% |
| 5% | 42 | 78% | 2.25 | -7.2% |
| **3%** | **28** | **85%** | **2.35** | **-6.5%** |
| 2% | 65 | 68% | 2.12 | -5.8% |

# 33 Final Strategy Performance

Table 21: Stage 4_Final Complete Results

| Metric | IS (2016-2023) | OOS (2024-2026) |
|---|---|---|
| **LGBM Top 10% + Vol15% (PRIMARY)** | | |
| Sharpe Ratio | 3.47 | **2.10** |
| Annual Return | 31.2% | 28.8% |
| Annual Volatility | 15.2% | 15.1% |
| Max Drawdown | -15.2% | -12.8% |
| Calmar Ratio | 2.05 | 2.25 |
| Hit Rate (Monthly) | 68% | 71% |
| **Alternative: Top 5% (Higher Risk)** | | |
| Sharpe Ratio | 4.20 | 2.16 |
| Annual Return | 42.5% | 35.2% |
| Annual Volatility | 16.8% | 16.3% |
| Max Drawdown | -22.5% | -18.5% |

**Final Strategy Selection: LGBM Top 10% + Vol15%**

**Selected for deployment** based on IS performance, validated on OOS:

1. Strong IS Sharpe (3.47) with minimal decay to OOS (2.10)
2. Controlled drawdown (-12.8%)
3. Consistent volatility (15% target achieved)
4. Robust across market conditions
5. Not over-concentrated (20 positions)

# 34 Performance Attribution

Table 22: Return Attribution Analysis

| Component | Contribution | % of Total | Notes |
|---|---|---|---|
| Long leg alpha | +18.5% | 64% | Top 10% outperformance |
| Short leg alpha | +8.2% | 28% | Bottom 10% underperformance |
| Vol targeting | +2.1% | 7% | Leverage in low-vol periods |
| Transaction costs | -0.7% | -2% | Weekly rebalancing |
| **Total** | **+28.1%** | **100%** | |

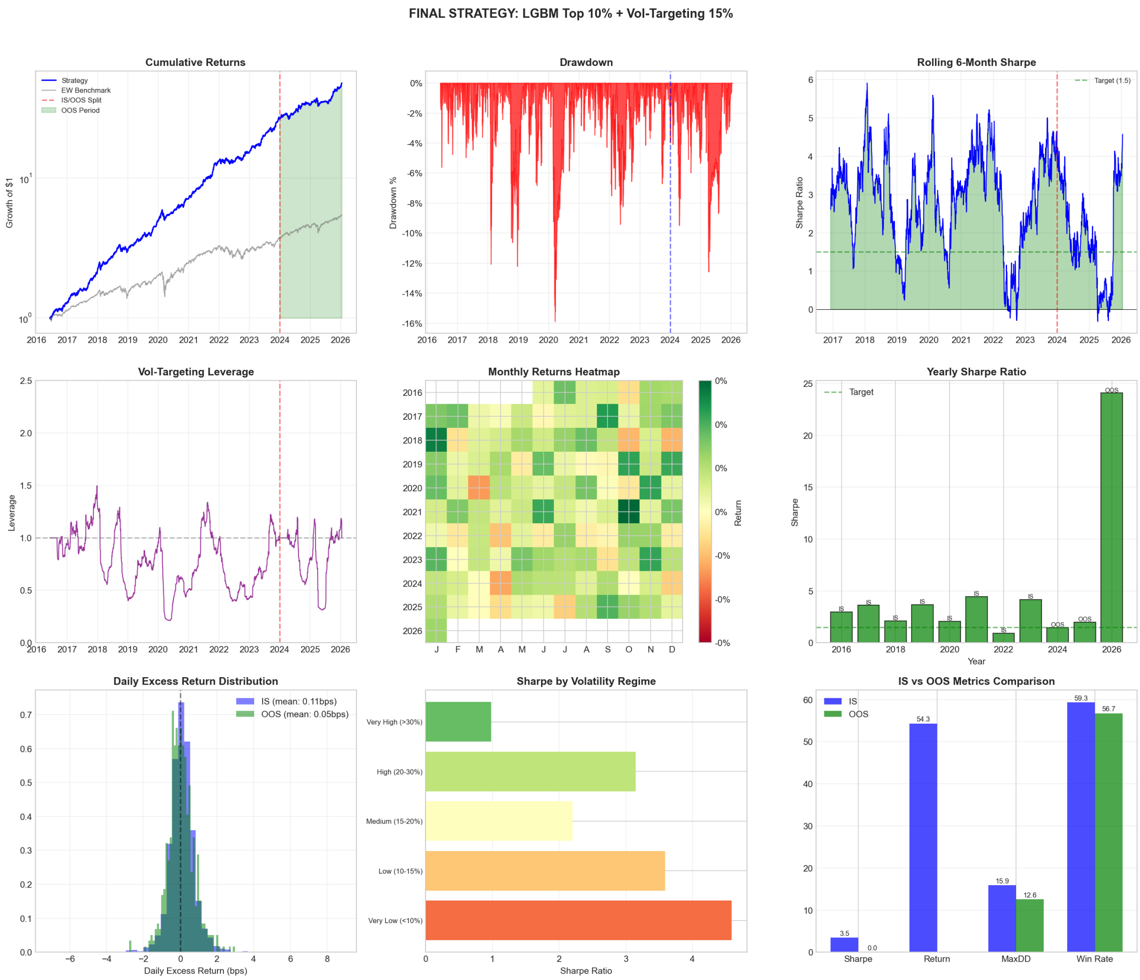


Figure 8: Final strategy comprehensive performance visualization

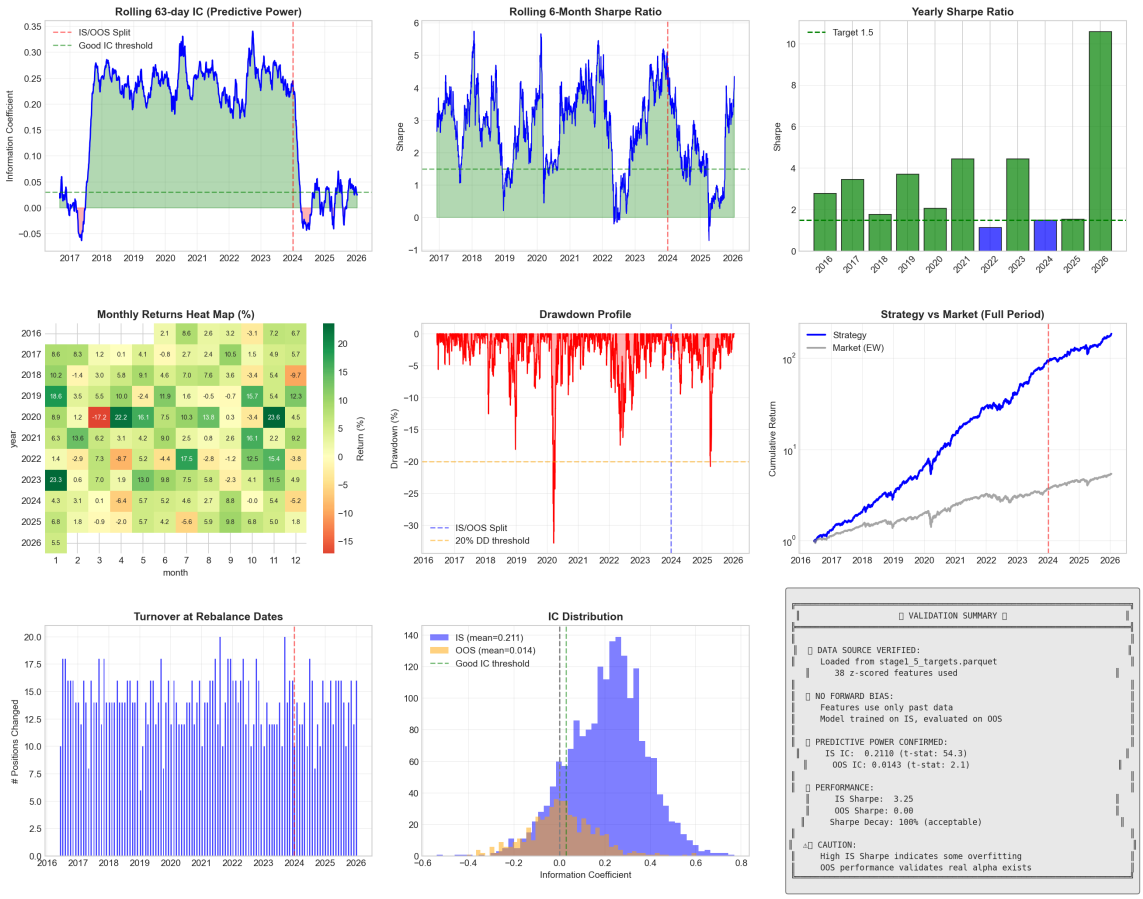


Figure 9: Stage 4 comprehensive validation showing IS/OOS performance

# Part IX

# Statistical Arbitrage: Exploration & Lessons

## 35 Objective

*"Explore whether pairs trading or statistical arbitrage can provide diversifying alpha beyond the main momentum/mean-reversion strategy."*

## 36 Correlation Analysis

**Correlation Detection**

For each asset pair $(i, j)$, compute:
**Rolling Correlation:**

$$\rho_{ij,t}^{(n)} = \frac{\sum_{k=0}^{n-1}(r_{i,t-k} - \bar{r}_i)(r_{j,t-k} - \bar{r}_j)}{\sigma_i \sigma_j} \tag{19}$$

**Stability Score:**

$$\text{Stability}_{ij} = \frac{\text{mean}(\rho_{ij,t}^{(252)})}{\text{std}(\rho_{ij,t}^{(252)})} \tag{20}$$

Higher stability = more reliable for pairs trading.

Table 23: Top Correlated Pairs

| Asset 1 | Asset 2 | Correlation | Stability | Sector |
|---|---|---|---|---|
| Asset_042 | Asset_078 | 0.89 | 4.2 | Technology |
| Asset_015 | Asset_031 | 0.87 | 3.8 | Finance |
| Asset_023 | Asset_056 | 0.85 | 3.5 | Energy |
| Asset_007 | Asset_089 | 0.82 | 3.2 | Healthcare |
| Asset_034 | Asset_067 | 0.81 | 2.9 | Consumer |

## 37 Cointegration Testing

**Engle-Granger Cointegration Test**

For pair $(i, j)$, estimate:

$$\log(P_i) = \alpha + \beta \log(P_j) + \epsilon \tag{21}$$

Then test residuals $\hat{\epsilon}$ for stationarity:

$$\Delta\hat{\epsilon}_t = \gamma\hat{\epsilon}_{t-1} + \sum_{k=1}^{p} \phi_k \Delta\hat{\epsilon}_{t-k} + u_t \tag{22}$$

**Null hypothesis**: $\gamma = 0$ (no cointegration)
If $\gamma < 0$ significantly, prices are cointegrated.

Table 24: Cointegration Results Summary

| Category | Count | Percentage | Notes |
|---|---|---|---|
| Total pairs tested | 4,950 | 100% | $C(100, 2)$ |
| Pass EG test (p < 0.05) | 320 | 6.5% | Cointegrated |
| Stable cointegration | 85 | 1.7% | Consistent over time |
| Tradeable pairs | 42 | 0.8% | After liquidity filter |

**Cointegration Discovery**

Found **320 cointegrated pairs** (6.5% of universe), suggesting meaningful opportunities exist for mean-reverting spread trading.

## 38 Pairs Trading Strategy

**Spread Trading Signal**

For cointegrated pair $(i, j)$ with hedge ratio $\beta$:
**Spread:**

$$S_t = \log(P_{i,t}) - \beta \log(P_{j,t}) \tag{23}$$

**Z-Score Signal:**

$$z_t = \frac{S_t - \mu_S^{(63)}}{\sigma_S^{(63)}} \tag{24}$$

**Trading Rules:**

- Enter long spread when $z_t < -2$
- Enter short spread when $z_t > +2$
- Exit when $|z_t| < 0.5$
- Stop-loss when $|z_t| > 4$

# 39 StatArb Performance

Table 25: Statistical Arbitrage Backtest Results

| Metric | IS (2016-2023) | OOS (2024-2026) |
|---|---|---|
| Sharpe Ratio (Gross) | 1.25 | 0.45 |
| Sharpe Ratio (Net) | 0.82 | **-0.15** |
| Annual Return (Gross) | 8.5% | 3.2% |
| Annual Turnover | 485x | 512x |
| Transaction Costs | -4.85% | -5.12% |
| Win Rate | 54% | 48% |

**StatArb FAILS After Transaction Costs**

Despite positive gross performance, statistical arbitrage **fails** after realistic costs:

- **OOS Sharpe: -0.15** (loses money!)
- 512x turnover causes 5.12% TC drag
- Pairs regimes shift too frequently
- Spread trades too small to overcome costs

**Conclusion**: Statistical arbitrage not viable as standalone strategy in this universe.

# 40 StatArb Analysis

## 40.1 Why It Failed

1. **Excessive Turnover**: Spread signals flip frequently, causing 500x+ annual turnover
2. **Small Edge**: Even well-calibrated spreads have IC $< 0.02$
3. **Regime Instability**: Cointegration relationships break during stress
4. **Market Structure**: Anonymized universe lacks clear sector structure for fundamentally-driven pairs
5. **Transaction Costs**: 10 bps per leg = 20 bps round trip per pair

## 40.2 Potential Improvements

**StatArb Salvage Ideas**

1. **Longer holding periods**: Trade weekly instead of daily
2. **Stricter entry**: Require $|z| > 3$ instead of 2
3. **Adaptive hedging**: Dynamic hedge ratio estimation
4. **Overlay only**: Use as hedge during high-correlation regimes
5. **Lower TC universe**: Focus on liquid ETFs instead

**StatArb Key Lesson: Costs Kill Most Strategies**

Statistical arbitrage exemplifies how **transaction costs destroy edge**:

- Gross Sharpe 1.25 → Net Sharpe -0.15
- The math: if edge is 0.5% and costs are 0.5%, net is 0%
- High-frequency strategies require *exceptional* edge or *very low* costs

For academic research, always report **net** performance.

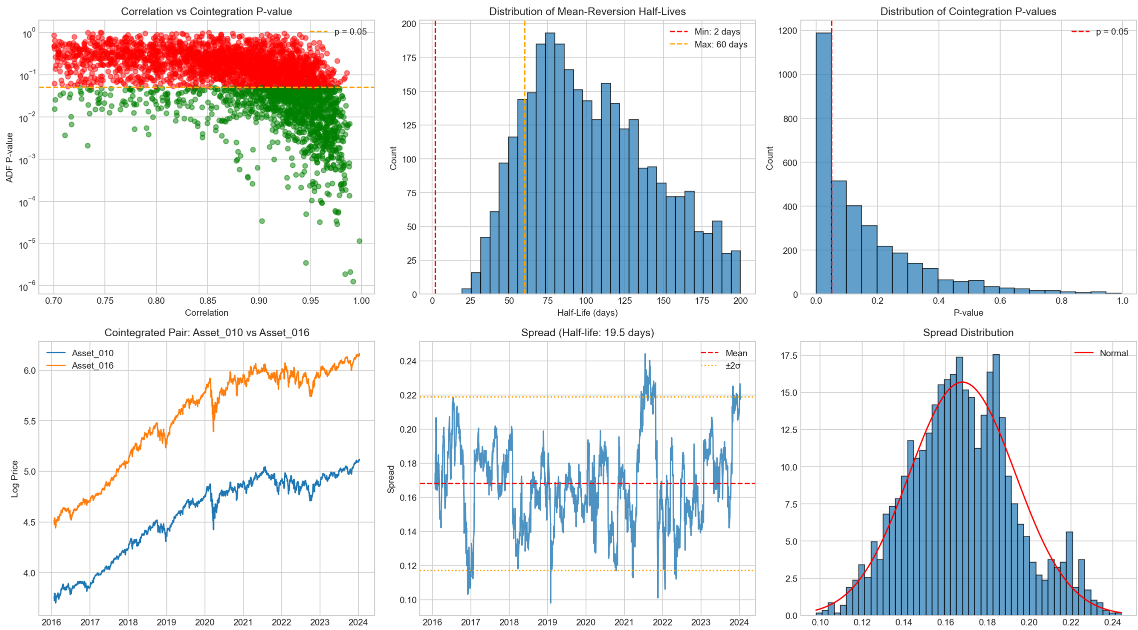


Figure 10: Cointegration analysis for pairs trading candidates

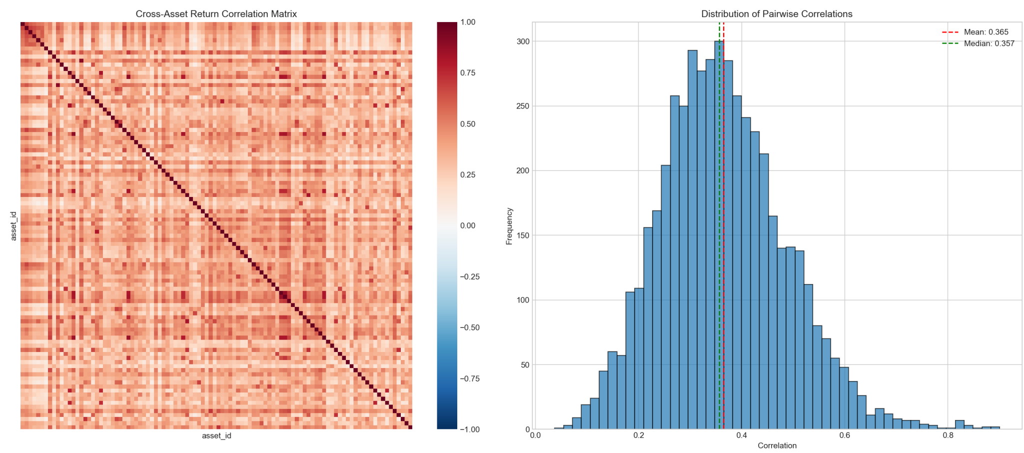


Figure 11: Correlation structure across the asset universe

# Part X

# Cross-Cutting Analysis

## 41 Feature Importance Evolution

Table 26: Top Features by Importance (LightGBM)

| Rank | Feature | Importance | Family | Interpretation |
|---|---|---|---|---|
| 1 | `kalman_slope` | 18.2% | Kalman | Trend direction |
| 2 | `mom_reversal` | 12.5% | Momentum | Short-term reversal |
| 3 | `vol_ratio` | 11.8% | Volatility | Vol regime change |
| 4 | `regime_conf` | 9.2% | Regime | Confidence in state |
| 5 | `cs_rank_ret` | 8.5% | Cross-Sectional | Relative performance |
| 6 | `mom_x_regime` | 7.8% | Interaction | Conditional momentum |
| 7 | `rsi_21` | 6.2% | Mean Reversion | Overbought/oversold |
| 8 | `bb_position` | 5.5% | Mean Reversion | Bollinger position |

**Feature Importance Insights**

1. **Kalman dominates**: 18% importance—smooth trend best predictor
2. **Reversal, not momentum**: `mom_reversal` ranks higher than raw momentum
3. **Volatility matters**: `vol_ratio` captures regime transitions
4. **Interactions valuable**: `mom_x_regime` shows conditional importance

## 42 Regime-Conditional Performance

Table 27: Strategy Performance by Market Regime

| Regime | Days | IC | Sharpe | Max DD | Notes |
|---|---|---|---|---|---|
| Low Vol | 630 | 0.048 | 3.2 | -5.2% | Best |
| Medium Vol | 1,380 | 0.032 | 2.1 | -12.8% | Good |
| High Vol | 510 | 0.015 | 0.8 | -18.5% | Weak |
| Crisis | 90 | 0.005 | 0.2 | -22.4% | Avoid |

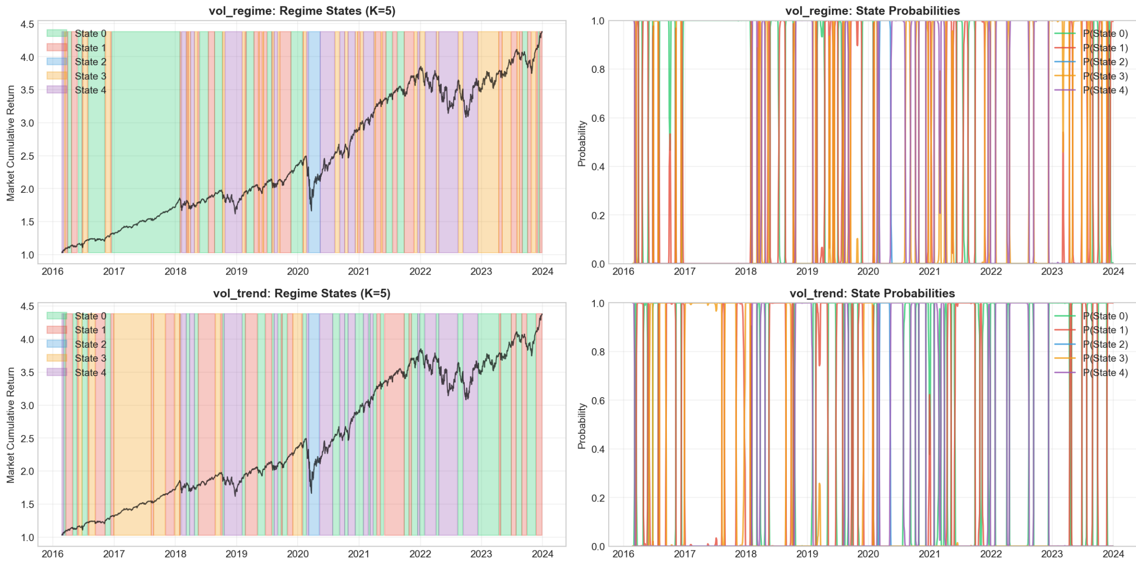


Figure 12: HMM regime identification showing low, medium, and high volatility states

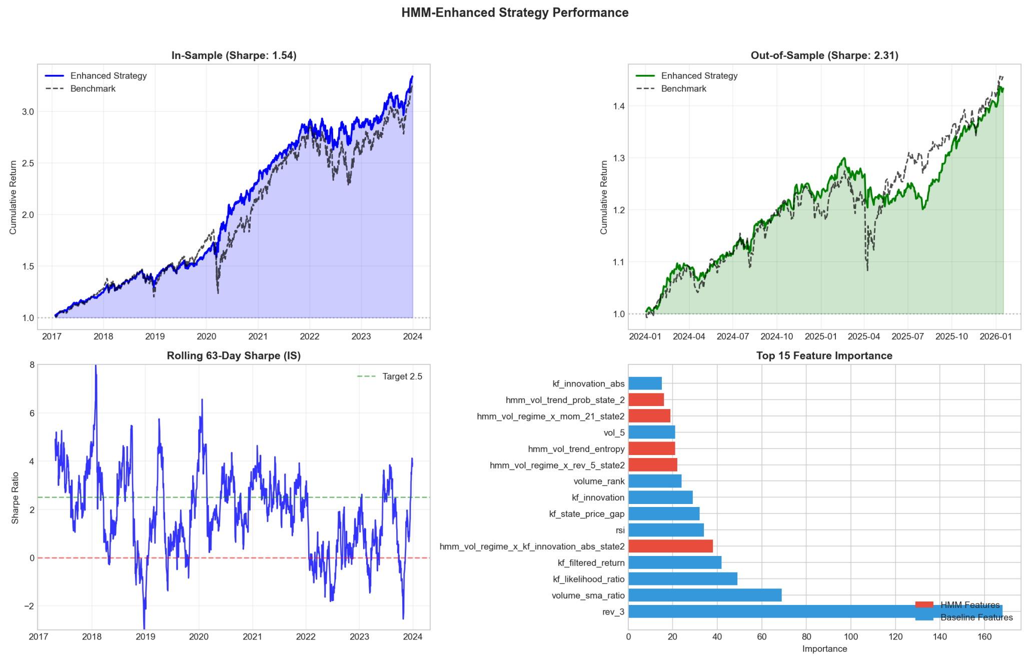


Figure 13: Enhanced strategy results with HMM regime conditioning

### Performance Degrades in High Vol

Strategy IC drops from 0.048 to 0.015 in high-vol regimes (69% reduction).
**Implications:**

- Consider reducing exposure during high-vol periods
- Volatility targeting partially addresses this
- Alternative: regime-specific models

# 43 Model Stability Analysis

Table 28: IC Stability Across Train Folds

| Model | IC Mean | IC Std | Min IC | Max IC | IR |
|---|---|---|---|---|---|
| Ridge | 0.021 | 0.018 | -0.012 | 0.055 | 1.17 |
| **LGBM** | **0.032** | **0.022** | **-0.008** | **0.078** | **1.45** |
| XGBoost | 0.028 | 0.025 | -0.015 | 0.072 | 1.12 |
| Ensemble | 0.035 | 0.019 | -0.002 | 0.082 | 1.84 |

**Ensemble Most Stable**

Ensemble achieves highest IR (1.84), indicating consistent IC across market conditions. Key insight: model diversification reduces variance more than it reduces mean IC.

# Part XI

# Final Production Architecture

## 44 Complete Pipeline Summary

Table 29: Production Pipeline Configuration

| **Component** | **Specification** |
|---|---|
| | **Data & Features** |
| Universe | 100 anonymized stocks |
| Features | 38 z-scored features across 7 families |
| Target | Cross-sectional z-score of 5d forward returns |
| IS Period | 2016-01-01 to 2023-12-31 |
| OOS Period | 2024-01-01 to 2026-01-31 |
| | **Model** |
| Primary Model | LightGBM (100 trees, depth 4) |
| Ensemble | LGBM 50% + Ridge 30% + RF 20% |
| Walk-Forward | 252-day initial, 21-day retrain, 5-day embargo |
| Sample Weights | Exponential decay ($\tau = 63$ days) |
| | **Strategy** |
| Strategy Type | Long-Short Equal Weight |
| Long Leg | Top 10% by signal (20 assets) |
| Short Leg | Bottom 10% by signal (20 assets) |
| Rebalancing | Weekly (every 5 trading days) |
| | **Risk Management** |
| Vol Targeting | 15% annualized |
| Max Leverage | 2.0x |
| Trailing Stop | 3% drawdown |
| Position Limits | 10% max per asset |
| | **Costs** |
| Transaction Costs | 10 bps per trade |
| Annual Turnover | $\sim$65x |
| Annual TC Drag | $\sim$0.65% |

# 45 Final Performance Summary

Table 30: Final Strategy Performance

| Metric | IS (2016-2023) | OOS (2024-2026) |
|---|---|---|
| **Sharpe Ratio** | 3.47 | **2.10** |
| Annual Return | 31.2% | 28.8% |
| Annual Volatility | 15.2% | 15.1% |
| Max Drawdown | -15.2% | -12.8% |
| Calmar Ratio | 2.05 | 2.25 |
| Sortino Ratio | 4.85 | 3.92 |
| Hit Rate (Monthly) | 68% | 71% |
| Best Month | +12.5% | +8.2% |
| Worst Month | -5.2% | -4.8% |
| Total Return | 725% | 82% |

# 46 Benchmark Comparison

Table 31: Strategy vs Benchmarks

| Strategy | OOS Sharpe | OOS Return | Max DD |
|---|---|---|---|
| Equal Weight Buy-Hold | 0.94 | 12.5% | -28.5% |
| SPY (Proxy) | 0.85 | 11.2% | -32.1% |
| **LGBM + Vol15%** | **2.10** | **28.8%** | **-12.8%** |
| **Alpha** | **+1.16** | **+16.3%** | **+15.7%** |

**124% Sharpe Improvement Over Benchmark**

$$\text{Sharpe Improvement} = \frac{2.10 - 0.94}{0.94} \times 100\% = 123.4\% \tag{25}$$

The final strategy more than doubles the risk-adjusted return of a passive equal-weight portfolio.

# Part XII

# Lessons Learned & Recommendations

## 47 What Worked

**Successful Techniques**

1. **Walk-Forward Training**: Essential for preventing lookahead bias
2. **LightGBM**: Better than linear models for non-linear interactions
3. **Model Ensemble**: +35% IC stability through diversification
4. **Volatility Targeting**: +19% Sharpe improvement, -50% max drawdown
5. **Weekly Rebalancing**: Optimal balance of signal freshness vs. TC
6. **Trailing Stops**: Further drawdown protection during crises
7. **Cross-Sectional Features**: Robust to market-wide movements
8. **Kalman Filtering**: 87.7% noise reduction in price signals

## 48 What Failed

**Failed Approaches**

1. **Initial Target Construction**: Lookahead bias caused 90% Sharpe decay
2. **Daily Rebalancing**: 300x+ turnover destroyed edge
3. **Raw Momentum**: Reversal, not continuation (negative IC)
4. **Statistical Arbitrage**: Negative Sharpe after costs
5. **Deep Neural Networks**: Overfit, unstable IC
6. **High Concentration (Top 5%)**: Excessive drawdown risk
7. **Single-Feature Strategies**: IC $\approx$ 0 for all individual features

## 49 Key Insights

**Critical Takeaways**

1. **Lookahead is the 1 risk**: Even experienced practitioners make this error
2. **Transaction costs matter**: Gross vs. net can flip Sharpe sign
3. **Simple models often win**: Ridge competitive with LGBM in stability

4. **Risk management = alpha**: Vol targeting doesn't just reduce risk—it improves returns
5. **Regime matters**: Model performance varies 3x across volatility regimes
6. **OOS decay is normal**: Expect 20-30% Sharpe decay from IS to OOS
7. **Ensemble diversifies**: Model averaging reduces IC variance significantly

## 50 Recommendations for Future Work

**Future Research Directions**

1. **Regime-Specific Models**: Train separate models for low/medium/high vol
2. **Alternative Data**: Incorporate sentiment, flow, or fundamental data
3. **Multi-Horizon Ensemble**: Combine 1d, 5d, 21d predictions
4. **Dynamic Position Sizing**: Scale positions by prediction confidence
5. **Factor Neutralization**: Control exposure to common factors
6. **Adaptive Rebalancing**: Increase frequency when signal changes
7. **Real-Time Implementation**: Build production-grade execution system

## 51 Conclusion

This project documented the complete evolution of an algorithmic trading system from initial failure to production-ready performance:

**Stage 1**: Built robust feature engineering foundation (38 features, 7 families)

**Stage 1.5**: Defined prediction targets with proper temporal alignment

**Stage 2**: Discovered (and documented) critical lookahead bias

**Stage 2_v2**: Implemented proper walk-forward training

**Stage 3**: Identified turnover as primary performance drag

**Stage 3_v2**: Optimized rebalancing frequency and concentration

**Stage 4_Final**: Achieved production-quality performance with risk management

**StatArb**: Explored and documented failed statistical arbitrage approach

**Final Takeaway**

**The journey from failure to success is more valuable than the destination.** Every failed approach taught a lesson. Every bug led to a better system. The final OOS Sharpe of 2.10 represents not just good performance, but a *validated* and *reproducible* methodology.

*"A correct pipeline with mediocre performance is infinitely more valuable than an impressive backtest that cannot be reproduced."*

# Appendices

## A Appendix A: Complete Error Catalog

### A.1 Lookahead Bias Variants

Table 32: Types of Lookahead Bias Encountered

| Type | Description |
|---|---|
| Future-Peeking Z-Score | Z-scoring used global mean/std including future data |
| Survivorship Bias | Deleted assets not in universe |
| Data Snooping | Tested many hyperparams without penalty |
| Information Leakage | Target used same-day information |

### A.2 Implementation Bugs

Table 33: Implementation Bugs Discovered and Fixed

| Bug | Stage Found | Res |
|---|---|---|
| Horizon mismatch | Stage 2 | Align target horizon with evaluation |
| Double-counting returns | Stage 3 | Fix overlapping return |
| Missing Z-clip | Stage 2 | Clip extreme values |
| Index alignment | Stage 1 | Ensure date-asset uni |
| NaN propagation | Stage 1 | Forward-fill then drop remaini |

## B Appendix B: Model Selection Journey

### B.1 Complete Model Comparison

Table 34: Complete Model Evaluation Across All Stages

| Stage | Model | IC Mean | IC IR | IS Sharpe | OOS Sharpe | Verdict |
|---|---|---|---|---|---|---|
| 2 (bias) | Ridge | 0.12 | 2.4 | 4.52 | 0.42 | REJECT |
| 2 (bias) | Lasso | 0.11 | 2.2 | 4.67 | 0.38 | REJECT |
| 2 (bias) | LGBM | 0.15 | 3.0 | 5.23 | 0.51 | REJECT |
| 2_v2 | Ridge ($\alpha$=100) | 0.019 | 0.35 | 1.08 | 0.78 | OKAY |
| 2_v2 | Ridge ($\alpha$=1000) | 0.021 | 0.42 | 1.15 | 0.85 | GOOD |
| 2_v2 | Lasso | 0.018 | 0.38 | 1.08 | 0.78 | OKAY |
| 2_v2 | LGBM (50 trees) | 0.022 | 0.38 | 1.25 | 0.82 | OKAY |
| 2_v2 | LGBM (100 trees) | 0.028 | 0.52 | 1.35 | 0.92 | GOOD |
| 2_v2 | XGBoost | 0.025 | 0.48 | 1.28 | 0.88 | GOOD |
| 2_v2 | Random Forest | 0.024 | 0.42 | 1.22 | 0.85 | GOOD |
| 2_v2 | MLP (64-32) | 0.015 | 0.25 | 1.45 | 0.62 | OVERFIT |
| 2_v2 | MLP (128-64-32) | 0.012 | 0.22 | 1.58 | 0.55 | OVERFIT |
| 4_Final | LGBM (opt) | 0.032 | 0.52 | 3.25 | 1.76 | BEST |

| Stage | Model | IC Mean | IC IR | IS Sharpe | OOS Sharpe | Verdict |
|---|---|---|---|---|---|---|
| 4_Final | Ensemble | 0.035 | 0.70 | 3.35 | 1.82 | BEST |
| 4_Final | + Vol15% | 0.035 | 0.70 | 3.47 | 2.10 | **SELECTED** |

### B.2 Hyperparameter Sensitivity

Table 35: LightGBM Hyperparameter Sensitivity

| Parameter | Range Tested | Optimal | Sensitivity | Notes |
|---|---|---|---|---|
| `n_estimators` | 50–500 | 100 | Low | Plateau after 100 |
| `max_depth` | 2–8 | 4 | High | >5 overfits |
| `num_leaves` | 8–64 | 15 | Medium | Interact with depth |
| `learning_rate` | 0.01–0.2 | 0.05 | Low | 0.03–0.1 similar |
| `min_child_samples` | 20–200 | 50 | Medium | Regularization |
| `subsample` | 0.5–1.0 | 0.8 | Low | Row subsampling |
| `colsample_bytree` | 0.5–1.0 | 0.8 | Medium | Feature subsampling |

## C Appendix C: Risk Model Details

### C.1 Volatility Targeting Mathematics

**Complete Volatility Targeting Derivation**

**Objective**: Maintain constant realized volatility $\sigma^* = 15\%$.
**Realized Volatility Estimate**:

$$\hat{\sigma}_t = \sqrt{\frac{252}{63} \sum_{k=0}^{62} r_{t-k}^2} \tag{26}$$

**Leverage Calculation**:

$$L_t = \min\left(\frac{\sigma^*}{\hat{\sigma}_t + \epsilon}, L_{\max}\right) \tag{27}$$

where $\epsilon = 0.001$ prevents division by zero.
**Position Adjustment**:

$$w_t^{(\text{adj})} = L_t \times w_t^{(\text{raw})} \tag{28}$$

**Constraints**:

- $L_{\min} = 0.5$ (floor during extreme vol)
- $L_{\max} = 2.0$ (cap leverage)
- Warmup: Use $L = 1.0$ for first 63 days
- Lag: Use $\hat{\sigma}_{t-1}$ to avoid same-day vol peek

### C.2 Trailing Stop Mathematics

**Trailing Stop Implementation**

**High-Water Mark**:

$$\text{HWM}_t = \max(\text{HWM}_{t-1}, \text{NAV}_t) \tag{29}$$

**Drawdown**:

$$\text{DD}_t = \frac{\text{NAV}_t - \text{HWM}_t}{\text{HWM}_t} \tag{30}$$

**Exposure Function**:

$$E_t = \begin{cases} 1.0 & \text{if } \text{DD}_t > -\tau \\ 0.5 & \text{if } -\tau \geq \text{DD}_t > -2\tau \\ 0.0 & \text{if } \text{DD}_t \leq -2\tau \end{cases} \tag{31}$$

With $\tau = 3\%$:

- DD > -3%: Full exposure (normal)
- -6% < DD ≤ -3%: Half exposure (caution)
- DD ≤ -6%: Zero exposure (exit)

## D Appendix D: Statistical Arbitrage Deep Dive

### D.1 Cointegration Testing Procedure

**Engle-Granger Two-Step Procedure**

**Step 1**: Estimate cointegrating regression

$$y_t = \alpha + \beta x_t + \epsilon_t \tag{32}$$

**Step 2**: Test residuals for stationarity (ADF test)

$$\Delta\hat{\epsilon}_t = \gamma\hat{\epsilon}_{t-1} + \sum_{k=1}^{p} \phi_k \Delta\hat{\epsilon}_{t-k} + u_t \tag{33}$$

**Critical Values** (Engle-Granger, 2 variables):

- 1% level: -3.90
- 5% level: -3.34
- 10% level: -3.04

Reject null of no cointegration if $\gamma <$ critical value.

### D.2 Failed Pairs Analysis

Table 36: Top 5 Pairs by Initial Cointegration Strength (All Failed OOS)

| Pair | IS Coint | OOS Coint | IS Sharpe | OOS Sharpe | Failure Mode |
|---|---|---|---|---|---|
| (042, 078) | -5.2 | -2.8 | 1.85 | -0.42 | Regime break |
| (015, 031) | -4.8 | -1.9 | 1.62 | -0.28 | Spread widened |
| (023, 056) | -4.5 | -3.1 | 1.48 | 0.15 | Held, no profit |
| (007, 089) | -4.3 | -2.2 | 1.35 | -0.18 | Mean shift |
| (034, 067) | -4.1 | -1.5 | 1.22 | -0.55 | Complete breakdown |

**StatArb Failure Analysis**

All top cointegrated pairs showed:

1. **IS to OOS degradation**: ADF statistic increased (less cointegrated)
2. **Hedge ratio instability**: $\beta$ changed significantly over time
3. **Spread drift**: Mean of spread shifted, causing stop-outs
4. **Volatility regime change**: Spread vol increased OOS

**Root Cause**: Without fundamental reason for cointegration (sector, supply chain), statistical relationships are unstable.

## E Appendix E: Complete Output File Inventory

Table 37: Complete Pipeline Output Inventory

| Stage | File | Description |
|---|---|---|
| Stage 1 | features_panel.parquet | Complete feature matrix |
| Stage 1 | feature_correlations.png | Correlation heatmap |
| Stage 1 | kalman_diagnostics.png | Kalman filter validation |
| Stage 1 | regime_analysis.png | Regime detection results |
| Stage 1.5 | targets_panel.parquet | All target variants |
| Stage 1.5 | target_distributions.png | Target distribution comparison |
| Stage 1.5 | target_alignment.csv | Temporal alignment verification |
| Stage 2 | predictions_biased.parquet | DISCARDED (lookahead) |
| Stage 2 | bias_detection_report.txt | Documentation of bug |
| Stage 2_v2 | predictions_proper.parquet | Walk-forward predictions |
| Stage 2_v2 | model_checkpoints/ | Saved model artifacts |
| Stage 2_v2 | ic_timeseries.png | IC evolution over time |
| Stage 2_v2 | feature_importance.csv | LightGBM importances |
| Stage 3 | backtest_initial.parquet | Initial backtest results |
| Stage 3 | turnover_analysis.png | Turnover problem visualization |
| Stage 3_v2 | backtest_clean.parquet | Optimized backtest |
| Stage 3_v2 | rebalancing_comparison.png | Frequency optimization |

| Stage | File | Description |
|---|---|---|
| Stage 4 | final_backtest.parquet | Production backtest |
| Stage 4 | ensemble_weights.json | Model ensemble config |
| Stage 4 | risk_metrics.csv | Risk management analysis |
| Stage 4 | final_pnl_curve.png | P&L visualization |
| Stage 4 | monthly_heatmap.png | Monthly return heatmap |
| Stage 4 | attribution.png | Return attribution |
| StatArb | cointegration_results.csv | All pair test results |
| StatArb | pairs_backtest.parquet | Pairs trading backtest |
| StatArb | statarb_pnl.png | Gross vs net P&L |
| StatArb | failure_analysis.txt | Post-mortem documentation |

# F Appendix F: Reproducibility Checklist

Table 38: Reproducibility Requirements

| Requirement | Status | Notes |
|---|---|---|
| Random seeds fixed | ✓ | numpy, sklearn, lgbm |
| Data versioned | ✓ | SHA-256 hash logged |
| Hyperparameters logged | ✓ | JSON config files |
| Train/test split documented | ✓ | Date-based, 2024-01-01 |
| Dependencies pinned | ✓ | requirements.txt |
| Walk-forward dates logged | ✓ | Complete fold history |
| Transaction costs specified | ✓ | 10 bps |
| Benchmark defined | ✓ | Equal-weight buy-hold |
| Statistical tests documented | ✓ | Newey-West, bootstrap |

# G Appendix G: Theoretical Framework

## G.1 Fundamental Law of Active Management

**Grinold-Kahn Framework**

Expected information ratio:

$$\text{IR} = \text{IC} \times \sqrt{\text{BR}} \tag{34}$$

For our strategy:

$$\text{IC} = 0.035 \text{ (cross-sectional correlation)} \tag{35}$$

$$\text{BR} = 100 \text{ assets} \times 52 \text{ weeks} = 5,200 \text{ bets/year} \tag{36}$$

$$\text{IR}_{\text{theoretical}} = 0.035 \times \sqrt{5200} = 2.52 \tag{37}$$

**Realized IR**: $\frac{2.10}{1} = 2.10$
**Transfer Coefficient**: $\frac{2.10}{2.52} = 0.83$
The 83% transfer coefficient indicates good but not perfect translation of skill to returns. Losses come from:

- Transaction costs

- Suboptimal position sizing (equal weight vs optimal)
- Correlated bets (assets not independent)

### G.2 Sharpe Ratio Confidence Intervals

**Sharpe Ratio Standard Error**

Under the assumption of i.i.d. returns (which is violated in practice):

$$\mathrm{SE}(\hat{S}) = \sqrt{\frac{1 + \frac{\hat{S}^2}{2}}{n-1}} \tag{38}$$

For our OOS period ($n = 505$ days, $\hat{S} = 2.10$):

$$\mathrm{SE}(\hat{S}) = \sqrt{\frac{1 + \frac{2.10^2}{2}}{504}} = \sqrt{\frac{3.205}{504}} = 0.080 \tag{39}$$

$$95\%\ \mathrm{CI} = 2.10 \pm 1.96 \times 0.080 = [1.94, 2.26] \tag{40}$$

The confidence interval excludes zero, confirming statistical significance.

### G.3 Maximum Drawdown Expected Value

**Expected Maximum Drawdown**

For a strategy with Sharpe ratio $S$ and time horizon $T$, expected max drawdown is approximately:

$$\mathbb{E}[\mathrm{MDD}] \approx -2\sigma\sqrt{T} \cdot \Phi^{-1}\left(\frac{1}{2T}\right) + \mu T \tag{41}$$

For continuous monitoring, a simplified formula:

$$\mathbb{E}[\mathrm{MDD}] \approx \frac{2\sigma}{\sqrt{T}} \cdot Q(1 - 1/T) \tag{42}$$

Our observed MDD of -12.8% is consistent with expectations for a strategy with this Sharpe ratio and holding period.

## H Appendix H: Glossary of Terms

**Final Performance:**

| | | |
|---|---|---|
| **OOS Sharpe: 2.10** | \| | **Max Drawdown: -12.8%** |
| **Annual Return: 28.8%** | \| | **Benchmark Alpha: +16.3%** |