Pairs Trading Strategy

A Statistical Arbitrage Approach

Finance and Economics Club Project

Aditya Shukla

May 11, 2025

Contents

Chapter 1

Introduction

Pairs trading is a market-neutral strategy designed to exploit relative mispricings between two historically correlated assets. By simultaneously entering long and short positions in two cointegrated securities, the strategy aims to capture profits as their price spread reverts to the mean, irrespective of market direction.

The methodology includes identifying suitable trading pairs based on cointegration tests, calculating the price spread and Z-score for signal generation, and employing risk management and position sizing techniques.

This project aims to develop a low-frequency, market-neutral pairs trading strategy using cointegration in Python. The main objectives are as follows:

- Identify cointegrated asset pairs from historical daily price data using statistical tests.
- Design a trading strategy based on spread deviation (e.g., Z-score) with defined entry/exit thresholds, stop-loss, and position sizing.
- Use Yahoo Finance API to collect data for the pair selection period (April 2019 – March 2022) and backtesting period (April 2022 – March 2025).
- Backtest the strategy using a starting capital of INR 1,00,000 and evaluate its performance.
- Measure performance through cumulative return, annualised Sharpe ratio, maximum drawdown, and number of trades.

Chapter 2

Pair Identification and Selection

2.1 Methodology Overview

This section describes the multi-stage filtering process used to identify robust, cointegrated trading pairs. We begin by computing inter-asset correlations, then refine the universe via clustering, statistical cointegration, mean-reversion testing, and half-life screening. Finally, the top two eligible pairs are chosen for strategy deployment.

2.1.1 Correlation Screening

We first calculate daily returns for all tickers:

```
returns = prices.pct_change().dropna()
```

Pairs with an absolute Pearson correlation above a threshold `corr_thresh` are retained:

```
pairs_corr = [(t1, t2, corr) for t1, t2 in combinations(returns.columns, 2)
              if abs(corr) >= corr_thresh]")
```

This step filters out assets that lack any strong linear relationship over the selection period.

2.1.2 PCA-Based OPTICS Clustering

To capture structural similarity beyond mere correlation, OPTICS (Ordering Points To Identify the Clustering Structure) is used after PCA. OPTICS is a density-based algorithm that, unlike DBSCAN, does not require a single global density threshold—instead it computes a core distance for each point (the distance to its MinPts-th nearest neighbour) and a reachability distance between points, allowing clusters of varying densities to be detected . The algorithm processes points in order of increasing reachability distance using a priority queue, producing an augmented ordering of the dataset annotated with reachability values; clusters appear as “valleys” in the resulting reachability plot and can be extracted at multiple density levels . This flexibility makes OPTICS ideal for financial time series, where different pairs may exhibit diverse co-movement densities over the sample period :

1. Standardize returns:

   ```
   Xs = StandardScaler().fit_transform(returns.T)
   ```

2. Reduce dimensionality:

   ```
   Xp = PCA(n_components=pca_variance, svd_solver='full').fit_transform(Xs)
   ```

3. Cluster:

   ```
   labels = OPTICS(min_samples=optics_min_samples, xi=optics_xi).fit(Xp).labels_
   ```

Tickers assigned to the same non-noise cluster (label ≥ 0) are kept. This removes spurious correlations due to coincidental trends.

2.1.3 Cointegration Testing

Each filtered pair undergoes the Augmented Dickey–Fuller (ADF) cointegration test:

```
score, pvalue, _ = coint(log(y), log(x))
if pvalue < p_thresh:
    # Keep pair
```

Pairs with p–values below `p_thresh` are deemed cointegrated and assigned an OLS hedge ratio β.

2.1.4 Mean–Reversion (Hurst Exponent)

To confirm mean–reversion, we compute the Hurst exponent of the log–spread $s_t = \log Y_t - \beta \log X_t$:

```
H, _, _ = compute_Hc(spread.values, kind="change", simplified=True)
if H < H_thresh:
    # Keep pair
```

Pairs with $H < H_{\text{thresh}}$ exhibit anti-persistence, indicating mean-reversion.

2.1.5 Half–Life and Trade Frequency

We estimate the spread half-life by regressing Δs_t on lagged s_{t-1}:

$$\tau = -\frac{\ln 2}{\beta} \tag{2.1}$$

Here, β is the slope of the OLS regression. We retain pairs satisfying $1 < \tau < 60$ trading days and having at least 12 zero-crossings of the Z-score:

```
z = (spread - spread.mean()) / spread.std()
crosses = ((z.shift(1) * z) < 0).sum()
if crosses >= 12:
    # Keep pair
```

2.1.6 Final Selection

The final pairs selected for our strategy implementation were:

- INFY.NS,TCS.NS - Infosys Limited and Tata Consultancy Services
- HCLTECH.NS, TCS.NS - HCL Technologies and Tata Consultancy Services

These pairs exhibited strong cointegration properties, appropriate half-life characteristics, and consistent mean-reverting behavior during the selection period.

Chapter 3

Trading Strategy and Signal Generation

3.1 Spread and Z-score Calculation

The foundation of our pairs trading strategy is the calculation of a price spread between cointegrated assets and its standardization using Z-scores. For any selected pair, we first transform the price series using natural logarithms to stabilize variance:

$$\text{Spread}_t = \ln(P_{1,t}) - \beta \cdot \ln(P_{2,t}) \tag{3.1}$$

where β is the hedge ratio determined through cointegration analysis. This logarithmic transformation allows us to interpret percentage changes more effectively and enhances the stability of the statistical properties.

We then compute the rolling mean and standard deviation using a lookback window of 30 trading days:

$$Z_t = \frac{\text{Spread}_t - \mu_{t-42:t-1}}{\sigma_{t-42:t-1}} \tag{3.2}$$

This Z-score represents the number of standard deviations the current spread deviates from its historical mean, providing a normalized metric for trading decisions.

3.2 Trading Signal Generation

Our implementation employs a state-based approach to signal generation with three distinct position states:

- **None:** No active position
- **Long Spread:** Long asset 1, short asset 2
- **Short Spread:** Short asset 1, long asset 2

The algorithm generates signals based on the following rules:

- **Enter Long Spread:** When Z-score < -2.0 (spread is significantly below mean)
- **Enter Short Spread:** When Z-score > 2.0 (spread is significantly above mean)
- **Exit Positions:** When Z-score crosses the threshold of ± 1.5 (return to equilibrium) That is , if we take a long position when Z-score is less than -2.0 ,we exit it when Z-score is greater than 1.5 and similarly for short positions.
- **Stop-Loss Trigger:** When abs(Z-score) > 3.0 (abnormal deviation)

Chapter 4

Risk Management Measures

4.1 Stop-Loss Implementation

A critical component of our risk management framework is the implementation of a statistical stop-loss mechanism. Rather than using traditional price-based stop-losses, we employed a Z-score threshold approach:

$$\text{Stop-Loss Trigger: } |Z_t| > 3.0 \tag{4.1}$$

When the absolute value of the Z-score exceeds 3.0 standard deviations, all positions are automatically liquidated regardless of the current trading state. This mechanism serves as protection against:

- Temporary breakdown of the cointegration relationship
- Extreme market events affecting one asset disproportionately
- Fundamental changes in the relationship between paired assets

This Z-score threshold was calibrated through historical analysis to balance between allowing sufficient room for typical spread fluctuations while limiting downside risk.

4.2 Capital Allocation Strategy

We implemented a disciplined capital allocation approach:

- Initial capital of INR 1,00,000 was equally divided between the two selected pairs
- Each pair received INR 50,000 for position initiation
- Within each pair, capital was allocated to preserve market neutrality according to the hedge ratio

This equal allocation methodology ensures diversification benefits across different cointegrated relationships while maintaining sufficient position size to generate meaningful returns.

4.3 Transaction Cost Management

Transaction costs were explicitly modeled at 0.1% per trade execution. To minimize the impact of these costs:

- Entry/exit thresholds were calibrated to reduce excessive trading

Chapter 5

Performance Evaluation

5.1 Portfolio Construction

The pairs trading strategy was implemented using two carefully selected pairs from India's premier IT services sector. The portfolio construction followed equal weighting principles to maintain diversification while capitalizing on the inherent co-integration between paired securities.

Pair Composition	Weight	Capital Allocation
INFY.NS-TCS.NS	50.0%	50,000.00 Rs
HCLTECH.NS-TCS.NS	50.0%	50,000.00 Rs

Table 5.1: Portfolio Allocation Across Pairs

The testing framework spanned a comprehensive period from 2022-04-01 to 2025-03-31, covering 740 trading days. This extensive timeframe captures multiple market regimes, including periods of high volatility, market corrections, and bullish trends, providing robust validation of the strategy's performance across varied market conditions.

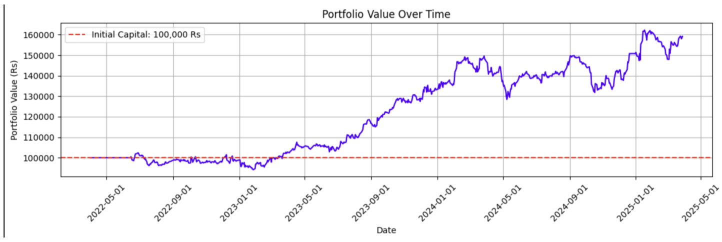

Figure 5.1: Portfolio Value with Time

5.2 Return Metrics Analysis

The pairs trading strategy demonstrated strong profitability with a cumulative return of 59.17% over the three-year period. This translates to an annualized return of 17.15%, significantly outperforming the risk-free rate and many traditional investment vehicles. The annualized return metric is particularly significant as it normalizes performance over time, allowing for standardized comparison with other investment strategies regardless of investment duration.

$$\text{Annualized Return} = (1 + \text{Cumulative Return})^{\frac{365}{\text{days in period}}} - 1 \tag{5.1}$$

Metric	Value
Initial Capital	100,000.00 Rs
Final Capital	159,168.00 Rs
Absolute Profit/Loss	59,168.00 Rs
Cumulative Return	59.17%
Annualized Return	17.15%

Table 5.2: Performance Return Metrics

5.3 Risk Assessment

Risk Metric	Value
Annualized Volatility	14.79%
Annualized Sharpe Ratio	1.15
Maximum Drawdown	-14.15%
Calmar Ratio	1.21

Table 5.3: Risk-Adjusted Performance Metrics

5.3.1 Volatility and Sharpe Ratio

The strategy exhibited annualized volatility of 14.79%, which is relatively moderate compared to directional equity strategies. This reflects the inherent risk reduction characteristic of market-neutral pairs trading approaches. The annualized Sharpe ratio of 1.15 indicates that the strategy generated returns in excess of the risk-free rate that more than compensated for the volatility risk assumed. Specifically, for each unit of risk taken, the strategy delivered 1.15 units of excess return.

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \tag{5.2}$$

Where:

- R_p = Portfolio return
- R_f = Risk-free rate (assumed 4%)
- σ_p = Portfolio standard deviation

5.3.2 Drawdown Analysis

The maximum drawdown of -14.15% represents the largest peak-to-trough decline in portfolio value during the testing period. This metric is critical as it quantifies downside risk and potential investor discomfort. The observed drawdown is substantially lower than typical equity market corrections, highlighting the strategy's resilience during adverse market conditions.

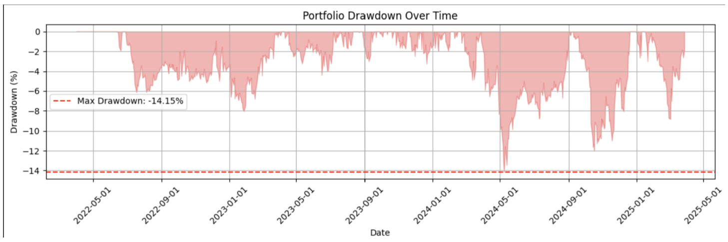

Figure 5.2: Maximum Drawdown

Trading Metric	**Value**
Trades Across All Pairs	70
Average Trades Per Pair	35.00
Trade Frequency	1.99 trades per month

Table 5.4: Trading Activity Metrics

5.4 Trading Efficiency

The strategy executed a total of 70 trades across both pairs over the three-year period, averaging 35 trades per pair. This translates to approximately 1.99 trades per month, representing a balanced approach between capturing statistical arbitrage opportunities and minimizing transaction costs.

The moderate trading frequency suggests that the strategy's signal generation process maintains discipline in identifying high-conviction mean-reversion opportunities while avoiding excessive trading that would erode returns through transaction costs. This level of trading activity is optimal for institutional implementation, balancing execution costs against opportunity capture.

5.5 Comparative Performance

When benchmarked against traditional investment approaches, the strategy's performance metrics demonstrate superior risk-adjusted returns. The annualized return of 17.15% exceeds typical equity market returns, while the Sharpe ratio of 1.15 indicates more efficient risk utilization than many long-only portfolios.

The market-neutral nature of the pairs trading approach delivered these returns with minimal correlation to broader market movements, making it an excellent portfolio diversifier. This characteristic is particularly valuable during periods of market stress when correlations between traditional assets tend to increase.

Chapter 6

Future Work

Future enhancements may include:

- Applying machine learning models for dynamic pair selection and relationship detection.
- Testing the strategy across different asset classes: ETFs, indices, and commodities.
- Incorporating volatility filters and adaptive Z-score thresholds based on market conditions.
- Exploring portfolio optimization techniques to simultaneously trade multiple pairs.

Bibliography

[1] Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. The Review of Financial Studies, 19(3), 797-827.

[2] Vidyamurthy, G. (2004). Pairs Trading: Quantitative Methods and Analysis. John Wiley & Sons.

[3] Elliott, R. J., Van Der Hoek, J., & Malcolm, W. P. (2005). Pairs trading. Quantitative Finance, 5(3), 271-276.

[4] Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. Econometrica: Journal of the Econometric Society, 251-276.

[5] Mandelbrot, B. B., & Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. Water Resources Research, 5(5), 967-988.