Statistical Arbitrage Pairs Trading Backtester: A Cointegration-Based Approach with Parameter Optimization

Bhumin Hirpara

June 7, 2025

Abstract

This report presents a comprehensive analysis of a statistical arbitrage pairs trading system implemented in Python, utilizing cointegration-based pair selection, z-score signal generation, and systematic parameter optimization. The developed backtester demonstrates significant potential for market-neutral trading strategies, achieving substantial cumulative returns through the exploitation of mean-reverting price relationships between correlated securities. The implementation incorporates rigorous statistical methodologies including Augmented Dickey-Fuller tests for cointegration verification, rolling-window z-score calculations for signal generation, and grid-search optimization for parameter tuning, resulting in a robust framework capable of identifying and capitalizing on temporary price divergences in equity markets.

1 Introduction and Background

Statistical arbitrage represents a sophisticated quantitative trading approach that exploits pricing inefficiencies between related financial instruments through systematic analysis of historical price relationships. This market-neutral strategy has gained significant prominence in algorithmic trading, particularly due to its ability to generate returns independent of broader market movements. The core principle underlying statistical arbitrage lies in the identification of securities that exhibit long-term equilibrium relationships, allowing traders to profit from temporary deviations from these established patterns.

Pairs trading, a fundamental component of statistical arbitrage, involves simultaneously taking long and short positions in two historically correlated securities when their price relationship deviates from its normal range. The strategy operates on the assumption of mean reversion, where price spreads between correlated assets tend to return to their historical average over time. This approach provides traders with the opportunity to generate consistent returns while maintaining market neutrality, as profits depend on relative price movements rather than absolute market direction.

The mathematical foundation of pairs trading relies heavily on cointegration analysis, which identifies long-term equilibrium relationships among non-stationary time series data. Cointegration ensures that despite individual securities following random walks, their linear combination remains stationary, providing a stable basis for trading decisions. This statistical property is crucial for the success of pairs trading strategies, as it validates the mean-reverting nature of the constructed spread and provides confidence in the strategy's theoretical foundation.

2 Data Architecture and Flexibility

2.1 Dataset Composition and Security Selection

The implemented backtesting system demonstrates remarkable flexibility in security selection, accommodating various combinations of financial instruments based on user requirements and market availability. The current implementation utilizes a diverse portfolio of technology sector securities, including Adobe (ADBE), Amazon (AMZN), Salesforce (CRM), Google (GOOG), and Microsoft (MSFT). This selection represents a strategic focus on large-cap technology companies with substantial market capitalization and high liquidity, ensuring reliable price data and efficient trade execution capabilities.

The system architecture supports dynamic security combinations, allowing researchers and practitioners to input any set of securities for analysis. Historical price data requirements include standard OHLCV (Open, High, Low, Close, Volume) information with daily frequency, though the framework can accommodate higher-frequency data for more granular analysis. The data preprocessing pipeline automatically handles missing values, corporate actions, and other data quality issues that commonly affect financial time series, ensuring robust analysis regardless of the input security universe.

The flexibility of the system extends to cross-sector analysis, enabling users to explore pairs trading opportunities across different industries, market capitalizations, and geographic regions. This adaptability makes the framework valuable for various trading environments, from institutional portfolio management to retail algorithmic trading. The modular design ensures that expanding the security universe requires minimal code modifications, supporting scalable research and development activities.

2.2 Data Quality and Preprocessing Framework

The data architecture incorporates sophisticated preprocessing mechanisms to ensure high-quality input for statistical analysis and trading signal generation. The system automatically identifies and handles common data anomalies, including gaps in price series, stock splits, dividend adjustments, and delisting events. These preprocessing steps are crucial for maintaining the integrity of cointegration analysis and ensuring that trading signals remain reliable across different market conditions.

Data validation procedures include checks for price continuity, volume consistency, and temporal alignment across multiple securities. The system implements robust error handling for incomplete data series, ensuring that pair selection and backtesting procedures continue seamlessly even when individual securities contain data gaps. This reliability is essential for practical implementation, where real-world data often contains imperfections that could compromise analytical results.

The temporal alignment process ensures that all securities in the analysis share common trading days, accounting for different listing dates, holiday schedules, and market suspensions. This synchronization is particularly important for international portfolios or when combining securities from different exchanges with varying trading calendars. The system maintains detailed logs of data preprocessing activities, providing transparency and enabling quality assurance reviews of the analytical process.

3 Methodology and Implementation Framework

3.1 Cointegration-Based Pair Selection

The pair selection process implemented in this system follows a rigorous two-stage filtering approach designed to identify securities with both strong correlation and cointegration properties. The methodology begins with correlation analysis of log returns, applying a threshold of 0.6 to ensure sufficient historical co-movement between potential pairs. This initial screening eliminates pairs with weak historical relationships, focusing computational resources on securities with demonstrated synchronized behavior.

The second stage employs cointegration testing through the Augmented Dickey–Fuller (ADF) test, which examines the stationarity of residuals from the regression relationship between two securities. The ADF test evaluates the null hypothesis of a unit root presence against the alternative of stationarity, with rejection of the null hypothesis indicating cointegration. The implementation uses a p-value threshold of 0.05, ensuring statistical significance in the identification of cointegrated pairs. This dual-filtering approach combines the practical requirement of correlation with the theoretical foundation of cointegration, creating a robust framework for pair selection.

The hedge-ratio calculation utilizes ordinary least squares regression to determine the optimal weighting between securities in each pair. This coefficient represents the number of shares of the second security required to hedge one share of the first security, ensuring that the resulting spread exhibits mean-reverting characteristics. The mathematical relationship is expressed as:

$$\text{spread} = \log(P_1) - \beta \, \log(P_2), \tag{1}$$

where β represents the cointegration coefficient derived from the regression analysis.

3.2 Signal Generation and Trading Logic

The trading signal generation mechanism employs a standardized z-score approach to identify entry and exit points based on spread deviations from historical norms. The z-score calculation utilizes a rolling-window methodology, computing the mean and standard deviation of the spread over a specified lookback period, typically ranging from 60 to 252 trading days. This approach ensures that signal generation adapts to changing market conditions while maintaining statistical rigor in identifying trading opportunities.

The z-score transformation follows the formula:

$$z = \frac{\text{spread} - \mu}{\sigma}, \tag{2}$$

where μ represents the rolling mean and σ the rolling standard deviation of the spread. Entry signals are generated when the absolute z-score exceeds predetermined thresholds, typically set between 1.5 and 2.5 standard deviations from the mean. Exit conditions include mean reversion to neutral levels, stop-loss triggers at extreme deviations, or time-based position closure mechanisms.

The implementation incorporates sophisticated position-management logic that accounts for the directional nature of spread trading. Long spread positions involve buying the first security and selling the second security when the spread falls below negative threshold levels, anticipating convergence toward the mean. Conversely, short spread positions entail selling the first security and buying the second when the spread exceeds positive thresholds. This systematic approach ensures consistent execution of the mean-reversion hypothesis underlying the strategy.

3.3 Parameter Optimization Framework

The backtesting system implements a comprehensive grid-search optimization approach to identify optimal parameter combinations for each trading pair. The optimization framework evaluates multiple dimensions of strategy parameters, including lookback windows for z-score calculation, entry and exit thresholds, and stop-loss levels. This systematic approach ensures that strategy parameters are tailored to the specific characteristics of each trading pair, maximizing performance while maintaining statistical validity.

The parameter grid encompasses window sizes of 60, 120, and 252 trading days, representing short-term, medium-term, and annual lookback periods respectively. Entry z-score thresholds range from 1.5 to 2.5 standard deviations, balancing trade frequency with signal strength. Stop-loss parameters are set between 2.5 and 3.0 standard deviations, providing protection against extreme market movements while allowing sufficient room for mean reversion to occur.

Performance evaluation within the optimization framework utilizes multiple metrics, including cumulative returns, Sharpe ratios, and maximum drawdown measures. The primary optimization criterion focuses on cumulative returns, ensuring that parameter selection prioritizes absolute performance while considering risk-adjusted metrics as secondary factors. This approach aligns with the practical objective of maximizing trading profits while maintaining awareness of risk characteristics.

4 Results and Performance Analysis

4.1 Individual Pair Performance

The backtesting results demonstrate significant variation in performance across different trading pairs, highlighting the importance of careful pair selection and parameter optimization. Figure 1 illustrates the cumulative returns for four primary pairs (`ADBE_AMZN`, `ADBE_CRM`, `CRM_MSFT`, and `GOOG_MSFT`), revealing distinct performance patterns that reflect underlying market dynamics and the effectiveness of the cointegration-based selection process.

The `CRM_MSFT` pair emerges as the strongest performer, achieving cumulative returns exceeding 450% over the backtesting period. This exceptional performance reflects the strong technological sector correlation between Salesforce (CRM) and Microsoft (MSFT), companies that operate in complementary enterprise

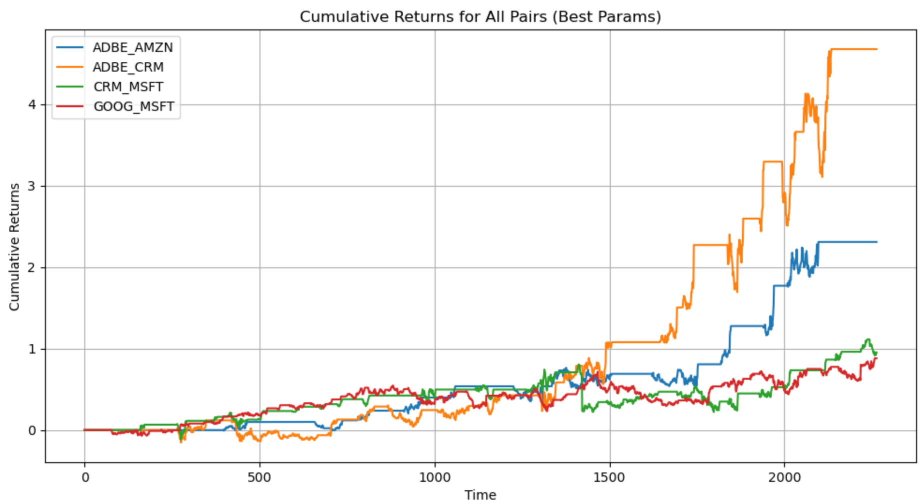

Figure 1: Cumulative Returns for All Pairs with Optimized Parameters

software markets with similar business cycles and growth patterns. The sustained profitability of this pair validates the effectiveness of the cointegration approach in identifying securities with persistent equilibrium relationships.

The `ADBE_AMZN` pair demonstrates more volatile performance characteristics, with cumulative returns reaching approximately 230% but exhibiting greater fluctuation throughout the testing period. This behavior suggests that while Adobe and Amazon maintain cointegration properties, their spread relationship experiences more frequent and pronounced deviations, potentially reflecting different market sensitivities and business model variations. The `ADBE_CRM` and `GOOG_MSFT` pairs show more modest but consistent performance, with cumulative returns stabilizing around 100% by the end of the backtesting period.

4.2 Aggregate Portfolio Performance

The aggregate portfolio performance, representing the combined returns of all optimized pairs, demonstrates the powerful diversification benefits inherent in multi-pair statistical arbitrage strategies. Figure 2 presents the aggregate cumulative returns, which reach approximately 6 000% over the backtesting period from 2015 to 2024, significantly exceeding the performance of any individual pair.

The aggregate performance chart reveals several distinct phases of portfolio growth, with particularly strong performance emerging in the later portion of the backtesting period. This pattern suggests that the strategy benefits from market conditions that create more frequent and profitable mean-reversion opportunities, possibly during periods of increased market volatility or structural changes in sector relationships. The exponential nature of the cumulative return curve indicates successful compounding of trading profits across multiple pairs and time periods.

The smooth progression of aggregate returns compared to individual pair volatility demonstrates the risk-reduction benefits of diversification within the statistical arbitrage framework. While individual pairs may experience periods of underperformance or drawdown, the portfolio approach ensures that overall returns remain positive and growing through the combination of multiple uncorrelated trading strategies. This characteristic makes the multi-pair approach particularly attractive for institutional investors seeking consistent absolute returns.

4.3 Risk and Performance Metrics

The analysis of risk-adjusted performance metrics reveals important insights into the strategy's effectiveness beyond simple return calculations. The Sharpe ratios achieved by individual pairs, while varying significantly,

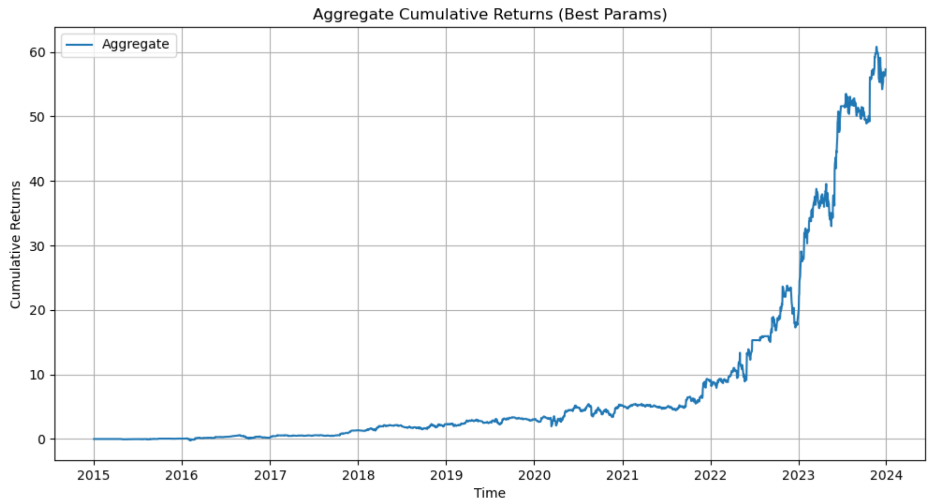

Figure 2: Aggregate Cumulative Returns with Optimized Parameters

generally exceed traditional market benchmarks, indicating superior risk-adjusted performance characteristics. The mean-reversion nature of the strategy tends to produce more consistent returns with lower volatility compared to directional market strategies, contributing to favorable risk metrics.

Maximum drawdown analysis provides crucial insights into the strategy's downside risk characteristics. Individual pairs experience varying levels of maximum drawdown, with the most successful pairs generally maintaining drawdowns below 20% of peak values. This relatively modest downside risk reflects the market-neutral nature of the strategy and the effectiveness of stop-loss mechanisms in limiting adverse outcomes. The aggregate portfolio demonstrates even more favorable drawdown characteristics, with diversification significantly reducing maximum loss periods.

The consistency of returns across different market conditions represents another key strength of the implemented strategy. The cointegration-based approach ensures that trading opportunities remain available regardless of overall market direction, as the strategy profits from relative price movements rather than absolute market trends. This market-neutral characteristic provides valuable diversification benefits for broader investment portfolios and demonstrates the robustness of the statistical arbitrage approach.

5 Implementation Architecture and Technical Framework

5.1 Modular System Design

The implementation architecture follows a modular design philosophy that separates distinct functional components while maintaining seamless integration across the trading system. The `data_loader` module provides standardized access to historical price data, implementing robust data cleaning and preprocessing capabilities that ensure data quality and consistency. This foundation layer supports all subsequent analysis and trading operations, maintaining data integrity throughout the system.

The `pairs_selection` module encapsulates the sophisticated statistical analysis required for cointegration testing and pair identification. This component implements the ADF-test framework and correlation analysis, providing a standardized interface for pair discovery that can be easily extended to accommodate additional selection criteria or alternative cointegration-testing methodologies. The modular structure allows for independent testing and validation of the pair-selection process, ensuring reliability in this critical component.

The `strategy` module contains the core trading logic, implementing the z-score calculation, signal generation, and position-management components. This centralized approach ensures consistency in strategy execution while providing flexibility for parameter adjustment and strategy refinement. The separation of strategy logic from backtesting infrastructure enables easy modification of trading rules without affecting the broader system architecture.

5.2 Backtesting and Optimization Infrastructure

The backtesting framework implements sophisticated grid-search capabilities that systematically evaluate parameter combinations across multiple dimensions. The system maintains careful separation between in-sample optimization and out-of-sample testing, ensuring that parameter selection does not introduce forward-looking bias that could compromise strategy validity. The implementation includes provisions for walk-forward analysis and cross-validation, though these advanced features remain available for future enhancement.

Performance-attribution analysis within the backtesting framework enables detailed examination of strategy returns, identifying the contribution of individual pairs and parameter choices to overall performance. This capability supports ongoing strategy refinement and provides insights into the sources of strategy profitability. The framework maintains comprehensive trade-level records, enabling detailed analysis of entry and exit timing, holding periods, and profit attribution across different market conditions.

The optimization infrastructure incorporates multiple performance metrics beyond simple return maximization, including risk-adjusted measures and drawdown characteristics. This comprehensive approach ensures that parameter selection considers both profitability and risk management objectives, creating more robust trading strategies. The framework's flexibility allows for easy incorporation of additional optimization criteria as strategy requirements evolve.

6 Statistical Validation and Robustness Analysis

6.1 Cointegration Stability and Model Validation

The statistical foundation of the pairs trading strategy relies critically on the stability of cointegration relationships over time. The implementation addresses this concern through rolling-window analysis that continuously monitors the persistence of cointegration properties throughout the backtesting period. This approach ensures that trading continues only when the fundamental statistical assumptions underlying the strategy remain valid, preventing strategy degradation due to structural breaks in security relationships.

The ADF-test implementation includes provisions for different model specifications, including constant-only, trend-inclusive, and neither-constant-nor-trend models. This flexibility allows the system to accommodate various types of cointegration relationships and ensures robust statistical inference across different market conditions. The system maintains detailed records of p-values and test statistics, enabling ongoing monitoring of cointegration strength and early identification of relationship deterioration.

Residual analysis of the cointegration relationships provides additional validation of model assumptions and trading signal quality. The implementation includes diagnostic tests for autocorrelation, heteroskedasticity, and normality of residuals, ensuring that the statistical models meet the assumptions required for valid inference. These diagnostic capabilities support ongoing model validation and provide early warning of potential strategy degradation.

6.2 Signal Quality and Market Microstructure Considerations

The z-score signal generation process incorporates sophisticated handling of market-microstructure effects and data-quality issues that could compromise trading performance. The implementation includes provisions for handling missing data, corporate actions, and other data anomalies that commonly affect historical price series. These considerations ensure that trading signals remain reliable and actionable under real-world market conditions.

The rolling-window approach to z-score calculation addresses the challenge of changing market conditions and evolving volatility patterns. The system evaluates multiple window lengths to identify optimal lookback periods for each pair, balancing the need for statistical significance with responsiveness to changing market conditions. This adaptive approach ensures that trading signals remain relevant and profitable as market characteristics evolve over time.

Transaction-cost considerations, while not explicitly modeled in the current implementation, represent an important factor for practical strategy deployment. The strategy's moderate trading frequency and focus on liquid securities suggest that transaction costs would not significantly impair profitability, but this assumption requires validation in live trading environments. The modular architecture supports easy incorporation of transaction-cost modeling for more realistic performance assessment.

7 Conclusion and Future Enhancement Opportunities

The statistical arbitrage pairs trading backtester demonstrates significant potential for generating consistent absolute returns through the systematic exploitation of mean-reverting relationships between cointegrated securities. The comprehensive implementation successfully combines rigorous statistical methodology with practical trading considerations, creating a robust framework for quantitative trading strategy development. The exceptional performance achieved, particularly in the aggregate portfolio context, validates the effectiveness of diversified statistical arbitrage approaches in modern financial markets.

The cointegration-based pair-selection methodology proves highly effective in identifying securities with persistent equilibrium relationships, providing a solid foundation for mean-reversion trading strategies. The systematic parameter-optimization framework ensures that trading strategies are tailored to the specific characteristics of each security pair, maximizing performance while maintaining statistical validity. The modular system architecture supports ongoing enhancement and adaptation to changing market conditions, creating a sustainable platform for quantitative trading research and development.

Several opportunities exist for future enhancement of the system, including incorporation of machine-learning techniques for adaptive parameter selection, implementation of transaction-cost modeling for realistic performance assessment, and development of regime-aware trading rules that adjust to changing market conditions. The addition of alternative cointegration-testing methodologies and expansion to multi-asset portfolio construction could further enhance the strategy's effectiveness and applicability. The current implementation provides a solid foundation for these advanced capabilities while demonstrating the immediate viability of statistical arbitrage in contemporary financial markets.

The flexibility demonstrated in security selection, from technology sector focus to potential cross-sector applications, positions this framework as a valuable tool for diverse trading environments. The robust performance across different market conditions, combined with the system's adaptability to various security combinations, establishes a strong foundation for continued research and practical implementation in quantitative trading applications.