UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Can Safety and Profitability Coexist?

Performance Analysis of Pairs Trading among S&P 500 Stocks

by

Mustafa Hussein

Supervisor: Adam Farago

Master's Thesis in Finance

Spring 2025

Graduate School, School of Business, Economics and Law, University of Gothenburg, Sweden

Abstract

This study examines the effectiveness of pairs trading strategies in U.S. equity markets from 2005 to 2024, with a particular focus on how trading volume affects performance. Using a comprehensive approach that analyzes 270 distinct parameter combinations, various weighting methodologies, and realistic transaction cost scenarios, the research demonstrates that volume is a fundamental driver of pairs trading success. High-volume pairs consistently outperform low-volume pairs across all tested configurations and optimal parameters yielding Sharpe ratios well above 2 in the testing period. Contrary to efficient market expectations, strategy performance improved in recent years (2015-2024) compared to the earlier period (2005-2014). Transaction cost analysis reveals that while the strategy remains successful under institutional-level cost structures, trading commission fees impact performance more significantly than borrowing costs. The findings contribute to both theoretical understanding of market efficiency and practical implementation of statistical arbitrage strategies, suggesting that pairs trading continues to offer attractive risk-adjusted returns for sophisticated investors despite its widespread documentation in financial literature.

Contents

1 Introduction

In today's financial landscape, investors navigate a fundamental trade-off. Safe fixed income investments like government bonds have delivered minimal returns of only about 1.54% on average over the last 20 years—typically insufficient to outpace inflation, as provided by French (2025). Conversely, the stock market has provided higher expected returns of around 9.1%, but with significant volatility and sensitivity to market conditions, potentially losing approximately 30% during crisis periods. As illustrated in Figure 1, market equities have delivered substantially higher cumulative returns compared to the risk-free rate, but with notable periods of significant volatility, particularly during the 2008-2009 financial crisis and the 2020 market downturn. This contrast leaves many investors, particularly risk-averse individuals seeking alternatives beyond traditional asset classes while avoiding complex derivatives, searching for a middle path with more balanced risk-return characteristics. This research investigates whether a reliable middle ground exists between these extremes, examining pairs trading as a potential alternative approach with distinct risk-return characteristics that might bridge this investment spectrum.

Figure 1: Cumulative returns of market equities and risk-free rate from 2005 to 2024

Pairs trading involves identifying two related securities—typically stocks—with a historically stable price relationship. This relationship often exists because the companies share similar business models, operate in the same market sector, face common economic drivers, or exhibit correlated stock movements due to fundamental market forces pushing their prices in similar directions. When this established relationship temporarily deviates, traders take opposite positions in both securities, going long on the relatively undervalued stock and short on the overvalued one. This approach aims to profit from price convergence, regardless of broader market movements. What distinguishes pairs trading is its potential to generate returns uncorrelated with the overall market. While traditional equity investments are exposed to market risk, properly constructed pair trades are designed to be beta-neutral, theoretically insulated from market-wide movements.

Technological advancements have democratized access to financial markets, with retail investors now having unprecedented access to trading platforms, data analytics tools, and low trading commission fees. This democratization makes previously exclusive strategies like pairs trading increasingly accessible to individual investors. As market volatility increases and traditional asset classes become more correlated during crises, market-neutral strategies grow more appealing.

Pairs trading, a strategy first popularized on Wall Street in the 1980s, has a rich history in institutional trading. However, its contemporary viability for today's investors remains an open question. This research addresses several interconnected questions about the strategy's current effectiveness and optimal implementation parameters.

The primary research question examines whether pairs trading remains effective in today's sophisticated market environment. Given the evolution of market efficiency, technological advances in trading, and increased market participation, a fundamental question arises: Is pairs trading still profitable in modern markets? The S&P 500 universe provides an ideal testing ground for this question, as it comprises highly liquid, well-established companies with extensive analyst coverage and institutional participation—conditions that theoretically should reduce market inefficiencies. By testing the strategy on these well-established companies, this research establishes a baseline for pairs trading performance in efficient markets.

Building upon this foundation, the research explores market microstructure effects on pairs trading performance. Under what specific market conditions does pairs trading generate superior returns? This question investigates the fundamental market dynamics that drive profitable opportunities. Specifically, the research examines whether the strategy performs better in highly traded securities where mispricings are quickly identified and corrected, or in less liquid securities where relative mispricings may persist longer before reverting to equilibrium relationships. This analysis provides insights into the underlying market mechanisms that create profitable pairs trading opportunities and helps identify optimal market segments for implementation.

Parameter robustness constitutes another critical area of inquiry. How sensitive are pairs trading returns to variations in implementation parameters? The practical application of pairs trading requires numerous parameter choices—data lookback periods, trading triggers, holding periods, and stoploss thresholds—each potentially affecting performance. By systematically varying these parameters and measuring their impact on returns, this research identifies stable parameter regions that produce consistent results across different market regimes, thus providing practical guidance for strategy implementation.

Closely related is the question of temporal stability. Has the effectiveness of pairs trading strategies evolved over time as markets have matured? By analyzing performance across different market cycles and regimes, this research examines whether the strategy's effectiveness has diminished as markets have become more efficient and algorithmic trading more prevalent. This temporal analysis helps investors understand whether pairs trading remains viable in contemporary markets or represents an arbitrage opportunity that has been gradually eliminated.

The research also explores portfolio construction approaches beyond simple pair selection. Does the weighting methodology significantly impact strategy performance? Given the diversity of potential pairs within the S&P 500 universe, optimal portfolio construction becomes crucial. This research compares equal-weighted, volatility-adjusted, and other weighting schemes to determine approaches that maximize risk-adjusted returns while managing concentration risk and exposure to specific sectors or factors.

The final research question addresses the practical implementation challenges by examining trading costs and implementation constraints. To what extent do realistic trading costs erode the theoretical profitability of pairs trading strategies? Academic studies often demonstrate profitability without fully accounting for real-world friction costs. By incorporating various cost structures into the analysis, this research provides a more realistic assessment of net returns and examines the minimum portfolio size needed for viable implementation in today's trading environment.

Despite its conceptual simplicity, successful pairs trading requires understanding when these strategies perform best. Their effectiveness varies based on market conditions, sector dynamics, and economic factors. Historical evidence suggests that pairs trading strategies often perform well during periods of market stress and volatility, when dislocations and subsequent returns to fundamental relationships between paired securities can generate substantial returns. However, during extreme market stress, even historically stable relationships between securities can break down, leading to substantial losses.

This research addresses these questions by systematically investigating the factors driving pairs trading success within S&P 500 stocks. Due to computational constraints, the analysis employs a targeted methodology: establishing a baseline using the entire S&P 500 universe, followed by more detailed examination of the volume extremes—highly traded versus thinly traded securities—to isolate the impact of trading liquidity on strategy performance.

The impact of this work extends beyond improving returns. By clearly explaining when pairs trading works and when it doesn't, this research aims to make this sophisticated strategy more understandable to regular investors. In an era where implementing trading strategies has become increasingly straightforward through fintech platforms, the key question shifts from "how to implement" to "when to implement" these approaches. This research seeks to answer the latter, helping investors recognize the specific market environments where pairs trading strategies are most likely to succeed and providing practical guidelines for effective implementation in contemporary markets.

2 Literature Review

Pairs trading[1] is a type of investment strategy that became popular on Wall Street in the 1980s. It's considered a form of statistical arbitrage - using statistical methods to identify mean-reverting behavior between securities and profit from temporary price divergences. Since then, researchers have implemented pairs trading in various ways and documented the effectiveness of different approaches across diverse market environments.

2.1 Foundational Research and Methodological Approaches

Gatev et al. (2006) established one of the first comprehensive frameworks for pairs trading. Their approach involved identifying stocks that historically moved together in price and trading on temporary divergences - buying the relatively undervalued stock and selling the relatively overvalued one. Their research spanning 40 years (1962-2002) demonstrated consistent profits, though declining in more recent periods. A key finding was the strategy's market-neutral characteristics, making it valuable for portfolio diversification. This early work provides important context for the first research question regarding whether pairs trading remains effective in today's sophisticated market environment.

Building on this foundation, Avellaneda and Lee (2010) developed more sophisticated mathematical methods for creating mean-reverting portfolios. Their study of U.S. stocks revealed performance variations across market conditions, with profits declining after 2007. They attributed this decline to increasing market efficiency and the rise of high-frequency trading - findings that directly inform the examination of temporal stability and whether pairs trading effectiveness has evolved over time.

The effectiveness of pairs trading fundamentally depends on the methodology used to identify potentially profitable pairs. Early implementations relied primarily on correlation metrics. Gatev et al. (2006) used a distance measure based on normalized price series. While intuitive, correlation-based approaches capture only linear relationships between returns rather than long-term equilibrium relationships between price levels. As B. Do and Faff (2012) noted, this often leads to pairs that fail to converge after divergence.

Cointegration, formalized by Engle and Granger (1987), represents a significant advancement in pairs trading methodology. Two stocks are cointegrated when a specific linear combination of their price series produces a stationary process, implying a stable mean-reverting relationship over time. This provides the theoretical foundation for convergence that pairs trading strategies rely upon.

Cointegration offers several advantages: stronger theoretical foundations in economic theory, statistical properties ensuring mean reversion, and more consistent performance across varying market conditions compared to correlation-based pairs, as demonstrated by Bowen and Hutchinson (2016). For the S&P 500 specifically, cointegration is well-suited given the high liquidity and complex interrelationships among large-cap stocks. As Jacobs and Weber (2015) observed, pairs trading strategies work best in markets with sufficient institutional participation to eventually correct mispricings - an observation relevant to

[1] While industry often labels portfolios of these strategies as 'statistical arbitrage,' this analysis uses 'pairs trading' to align with academic literature.

the research question examining market microstructure effects on strategy performance.

Despite its advantages, cointegration has limitations. The stationarity relationship may break down during regime changes, and H. X. Do and Brooks (2015) noted that cointegration relationships may weaken over time as markets become more efficient, requiring periodic recalibration. These limitations highlight the importance of the research question regarding parameter robustness and sensitivity to implementation choices.

2.2 Evolution of Research: Advanced Techniques and Applications

As technology has advanced, pairs trading methods have grown increasingly sophisticated. Figuerola-Ferretti et al. (2018) enhanced traditional approaches by analyzing price spread persistence in European markets, providing valuable information about mean reversion timing.

Machine learning has opened new possibilities for pairs trading. Flori and Regoli (2021) demonstrated how Long Short-Term Memory (LSTM) networks could improve results by detecting hidden patterns in market data. Their model predicted potential relationship reversals rather than simply providing direct trading signals. Taking a different approach, Han et al. (2023) employed unsupervised learning to group stocks based on company characteristics and price movements. Their strategies performed particularly well during crisis periods, with significantly smaller drawdowns than traditional methods. These findings speak to the research questions about temporal stability and performance across different market conditions.

While most research focuses on price relationships, some researchers have explored how company fundamentals affect pair selection. Hong and Hwang (2023) studied U.S. markets from 2001 to 2020 and found that pairs with similar fundamental characteristics performed better than those selected based only on price patterns. This approach reduced non-convergence risk, though they noted declining profits since 2003 even when accounting for fundamental similarities - a trend that reinforces the importance of the primary research question about the continued viability of pairs trading strategies.

Chen et al. (2019) investigated U.S. markets from 2000 to 2015 using different statistical methods to select pairs. They identified short-term price reversals and industry momentum as key contributors to pairs trading profits, demonstrating that multiple factors drive these strategies' success. This multi-factor perspective relates to the questions about market microstructure effects and optimal portfolio construction approaches.

2.3 Market Conditions and Implementation Challenges

Several researchers have questioned pairs trading's sustained profitability. B. Do and Faff (2012) examined how trading costs affect returns, finding that real-world costs significantly reduce academic study profits. They also observed declining convergence probability over time, suggesting increasing market efficiency. These findings directly inform the final research question addressing trading costs and implementation constraints.

Market conditions strongly influence success rates. Jacobs and Weber (2015) analyzed 34 international markets and discovered that company-specific news-driven divergences tend to produce lower returns than broader economic factor-driven divergences. Increased investor attention leads to reduced profits as heightened scrutiny causes faster price corrections. Their research highlighted how arbitrage limitations (high trading costs, short-selling restrictions) significantly impact performance - considerations essential to the examination of practical implementation challenges.

Many studies have noted that pairs trading effectiveness varies significantly with market conditions. Bowen and Hutchinson (2016) studied pairs trading in the UK stock market from 1992 to 2008 and found that cointegration-based strategies performed consistently, though results varied across different market environments. Particularly relevant to the current research questions, several studies show that well-designed pairs trading strategies often perform surprisingly well during market crises. Han et al. (2023) documented strong performance during both the 2007 financial crisis and 2020 pandemic market disruption, with maximum drawdowns limited to approximately 12%. This suggests that market disruptions may create enhanced opportunities for statistical arbitrage strategies - a research that the temporal stability analysis will test.

A relatively underexplored area, which this research specifically addresses, is the relationship between trading volume patterns and strategy performance. While many studies control for liquidity, few directly examine how volume dynamics affect convergence probability and profit potential. B. Do and Faff (2012) touched on this when examining implementation costs, noting that high-volume pairs often faced greater market impact costs but also demonstrated more reliable convergence. Market microstructure research by Jacobs and Weber (2015) hints that institutional trading patterns, often reflected in volume signatures, may provide signals about the nature of divergence events and their likelihood of correction. The present research extends this work by explicitly comparing strategy performance between highly traded and thinly traded securities within the S&P 500 universe.

Krauss (2017) provided a comprehensive review highlighting pairs trading's evolution and the need for adaptation as markets become more efficient. More recently, Sarmento and Horta (2020) explored high-frequency pairs trading using 5-minute data intervals with machine learning models. Their approach achieved better risk-adjusted returns than traditional methods, though they noted a trade-off between reducing drawdowns and overall profitability. These developments in research approaches demonstrate the necessity of the parameter robustness examination conducted within this study.

2.4 Research Gaps and Future Directions

While existing research provides valuable insights, several important questions remain unanswered that this research aims to address:

1. Most studies examine broad markets rather than focusing specifically on S&P 500 stocks, which represent the most liquid and widely-traded securities in the U.S. market. This research addresses this gap by establishing a baseline for pairs trading performance within this specific universe of

securities.

2. The relationship between pairs trading performance and specific market conditions is not fully understood, particularly how these strategies behave during transitions between calm and volatile market regimes. The temporal stability analysis will contribute to this understanding.

3. How stock-specific characteristics like trading volume affect pairs trading success has received limited attention, with mixed findings across different markets and time periods. The targeted methodology examining volume extremes directly addresses this research gap.

4. The practical implementation challenges, including parameter sensitivity and trading costs, have been acknowledged but not comprehensively analyzed for the S&P 500 universe specifically. The parameter robustness and implementation constraints analyses will provide insights for practitioners.

5. The optimal approach to portfolio construction for pairs trading strategies remains an open question, particularly regarding weighting methodologies. This research will compare various weighting schemes to identify approaches that maximize risk-adjusted returns.

By addressing these gaps through the specified research questions, this study aims to provide a more complete understanding of when and why pairs trading strategies succeed or fail in large-cap U.S. stocks. This knowledge would help risk-averse investors identify the most favorable conditions for implementing pairs trading strategies as a middle-ground approach between low-yielding fixed income investments and volatile equity markets.

3 Data

This study draws upon multiple data sources to ensure comprehensive and robust analysis of pairs trading strategies. The foundation of this research utilizes daily stock price data from the Center for Research in Security Prices CRSP (2025) specifically focusing on S&P 500 constituent stocks. The dataset has been carefully filtered to include only entries for periods during which companies were actual components of the S&P 500 index. This filtering approach ensures that the analysis remains focused on large-cap, liquid securities that would be practical for implementing pairs trading strategies in real-world conditions.

For market benchmarking and risk-adjusted performance evaluation, the study utilizes data from French's (2025) Data Library. This includes risk-free rate data (1-month Treasury bill rates) and market return data for calculating excess returns. This provides essential baseline measurements against which to compare the performance of pairs trading strategies in various market conditions.

3.1 Sample Period and Market Environments

The study covers a 20-year trading period from 2005 to 2024, intentionally selected to capture a wide range of market conditions and economic cycles. It includes the pre-crisis stability period from 2005 to 2007, the global financial crisis of 2008–2009, and the subsequent recovery phase from 2010 to 2015. The analysis also covers the late-stage bull market of 2016–2019, the COVID-19 market disruption in 2020, and the post-pandemic market environment from 2021 to 2024. This broad timeframe provides a robust foundation for evaluating the performance of pairs trading strategies across diverse and shifting market regimes.

The primary dataset begins with raw daily stock price data from the CRSP database. As part of the data preparation process, this data is cleaned and adjusted for stock splits and dividends to create continuous price series suitable for pairs trading analysis. This adjustment process is essential as it removes artificial price jumps that would otherwise distort the statistical relationships between potential stock pairs.

In addition to the adjusted price data, several other metrics are included. Trading volume data provided by CRSP (2025) delivers insights into liquidity conditions, which can significantly impact the execution quality and transaction costs of pair trades. Market capitalization helps in understanding the relative size of companies and potentially grouping pairs based on similar market cap ranges.

3.2 Data Processing and Visualization

Before analyzing the data, several preprocessing steps were applied to adjust stock prices properly. Survivorship bias was mitigated by only including stocks during their S&P 500 membership periods, acknowledging the dynamic nature of index composition while focusing on large, liquid companies. All price data has been adjusted for corporate actions including dividends, stock splits, and mergers to maintain continuous time series.

Extreme price movements exceeding daily thresholds of ±20% were examined individually to verify their accuracy and determine whether they represented genuine market movements or data errors. Trading days with missing price data were addressed using forward-fill methodology for temporary gaps of one or two days, while longer gaps resulted in the exclusion of those securities from pair formation during the affected periods.

For the market data from Kenneth French's library, the daily observations were aligned with the CRSP dataset, ensuring proper matching of dates and accounting for market holidays and weekend gaps.

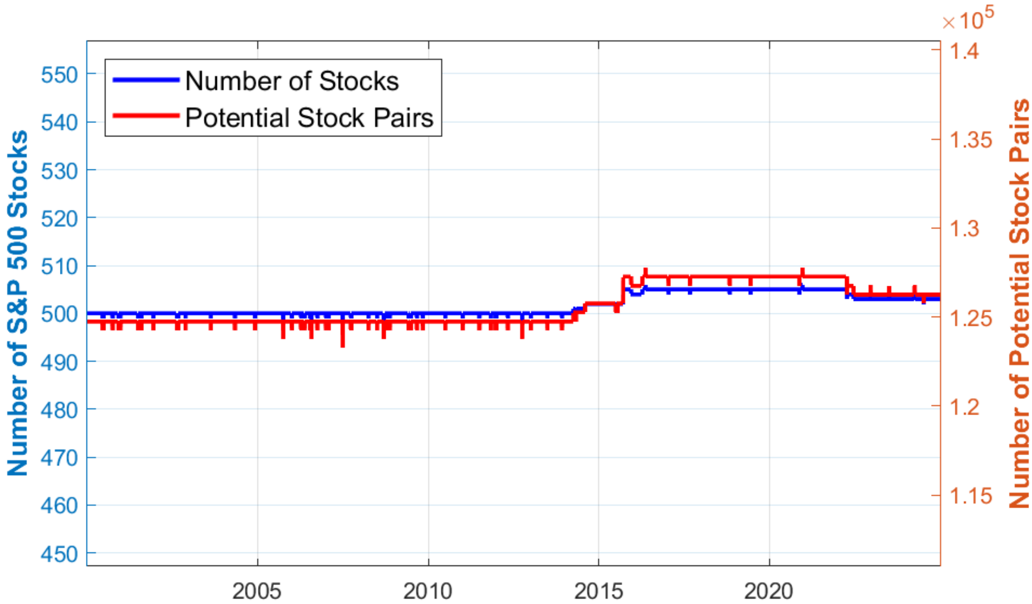


Figure 2: S&P 500 constituent count (left axis) and potential trading pairs (right axis), 2005-2024, showing index composition fluctuations that impact the available pool for pair selection.

To provide readers with a clear understanding of the dataset's scope and characteristics, several visualizations are presented. Figure 2 illustrates the dynamic nature of the dataset. The primary axis shows the number of S&P 500 stocks included in the analysis over the 20-year period, reflecting changes in index composition. The secondary axis displays the corresponding number of potential stock pairs (calculated as $\frac{n(n-1)}{2}$ where n is the number of stocks), highlighting the combinatorial growth of pairing possibilities. Having 500 stocks would theoretically lead to 124,750 possible combinations.

Interestingly, Figure 2 reveals that the S&P 500 index did not consistently maintain exactly 500 constituent stocks throughout the study period. During the years leading up to 2014, the actual number of constituents occasionally fell below 500, with notable decreases during 2006-2008 (coinciding with the financial crisis) due to bankruptcies, mergers, or companies failing to meet index inclusion requirements related to market capitalization or other criteria. After 2015, the number of constituent companies occasionally exceeded 500, reaching as high as 510 at certain points. These fluctuations in the number of companies directly affect the number of potential pairs available for analysis.

4 Methodology

The methodology follows a systematic process as illustrated in the flowchart in Figure 24 in the Appendix (section 8.1), beginning with comprehensive data collection and preprocessing, followed by pair identification through statistical testing, trade execution based on predetermined signals, and performance evaluation. The flowchart captures the complete lifecycle of the pairs trading strategy implementation. Starting with data collection from the S&P 500 universe, the process flows through multiple decision points and operational steps. After initial data preprocessing, the methodology applies cointegration testing to identify statistically related pairs. For qualified pairs, the process calculates and normalizes spreads to generate standardized z-scores. The trading logic then monitors these z-scores for entry signals (when absolute z-scores exceed the entry threshold of 2), initiating positions accordingly. Once positions are established, the strategy continuously monitors for exit conditions: either convergence to the mean (z-score within ± 0.5), triggering of stop-loss limits, or reaching the maximum holding period. This systematic approach ensures consistent implementation across hundreds of potential pairs.

This methodology addresses the research questions by examining how pairs trading performance varies across different market environments and trading volumes, while identifying optimal parameters and portfolio construction techniques.

4.1 Pairs Selection and Signal Generation

The foundation of successful pairs trading lies in identifying stocks with strong statistical relationships. The process begins by examining historical price data for potential stock pairs within the S&P 500 universe. For each potential pair, the analysis uses a 4-year lookback period of daily price data to establish reliable statistical relationships.

The first step involves running linear regression analysis between the price series of two stocks. This can be expressed in matrix form as:

$$P_1 = X\beta + \varepsilon \tag{1}$$

Where P_1 represents the price vector of the first stock, X is the design matrix $[1\ P_2]$ containing a column of ones for the intercept and a column with the price vector of the second stock, $\beta = [\alpha\ \gamma]'$ is the parameter vector containing the intercept (α) and hedge ratio (γ), and ε is the vector of residuals. The hedge ratio (γ) indicates the proportion of the second stock needed to hedge the position in the first stock.

The residual between the paired stocks is calculated as:

$$Z = P_1 - X\beta \tag{2}$$

Where Z is the vector of residuals representing the deviations from the fitted relationship. This parameter Z captures the degree of mispricing between the two stocks after accounting for their historical relationship through the parameter vector β. In a perfectly hedged position, these residuals should fluctuate around zero, with deviations representing temporary pricing inefficiencies that the strategy aims to exploit.

After calculating the spread, the Johansen cointegration test is applied to validate the long-term equilibrium relationship between stock prices. The Johansen test, developed by Søren Johansen (1988), extends the concept of cointegration to multivariate systems and provides a more robust framework than simpler methods like the Engle-Granger test. Cointegration implies that while two or more time series may individually follow random walks (non-stationary processes), a specific linear combination of them results in a stationary process that tends to revert to a mean value over time. The test is formalized through a vector error correction model:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta X_{t-i} + \varepsilon_t \tag{3}$$

Where $X_t = [P_{1,t}, P_{2,t}]'$ represents the vector of price series, Π captures the long-run relationship, and Γ_i represents short-term dynamics. The test examines the rank of matrix Π to determine if cointegration exists. Only pairs passing this test with a p-value below 0.05 are considered for trading, ensuring that the identified relationships are statistically significant and likely to exhibit true mean-reverting behavior rather than spurious correlations. This rigorous statistical validation is critical because pairs trading profitability fundamentally depends on the reliability of mean reversion in the spread. The Johansen test applied to pairs trading by researchers like Chan (2013) and Tadi and Davallou (2021), has become a standard approach in the field due to its ability to handle multiple variables and identify cointegrating relationships more effectively than earlier methods.

Once cointegrated pairs are identified, the residual vector Z is normalized to generate standardized scores:

$$z = \frac{Z - \mu_Z}{\sigma_Z} \tag{4}$$

where z is the vector of standardized scores, μ_Z is the mean of the residuals and σ_Z is the standard deviation of the residuals over the lookback period. This normalization creates standardized trade signals across different pairs, allowing for consistent application of entry and exit thresholds regardless of the absolute magnitude of price movements in the underlying securities.

Figure 3 illustrates this process using Chevron (CVX) and ExxonMobil (XOM) stocks as an example. The top panel shows how these oil industry peers generally move together but occasionally diverge. The bottom panel displays the calculated z-score with trading thresholds indicated by horizontal lines. When the z-score exceeds +2 standard deviations, the strategy signals to short CVX and long XOM, anticipating convergence. When it falls below -2 standard deviations, the strategy signals to long CVX and short XOM.

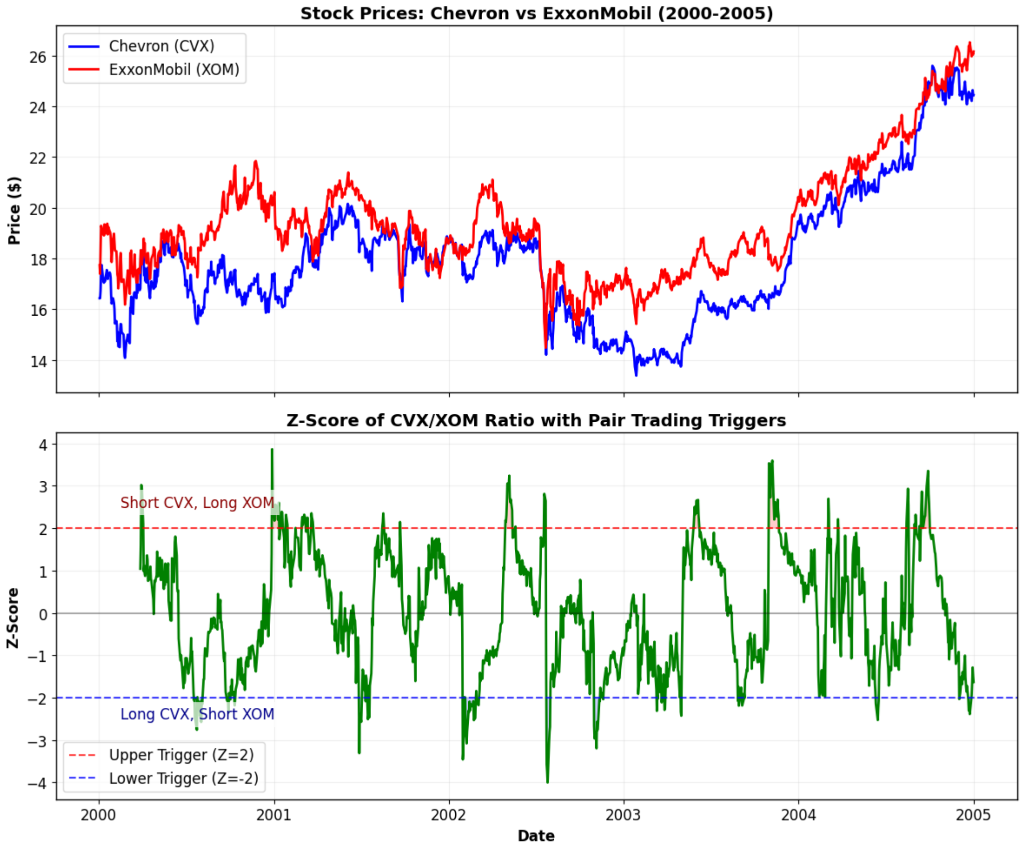


Figure 3: Pairs trading strategy illustration using Chevron (CVX) and ExxonMobil (XOM) stock prices and z-score triggers (2000-2005).

Once a position is established, it is monitored daily for exit signals. Two primary exit mechanisms are implemented. Convergence exit occurs when the z-score returns to within 0.5 standard deviations of the mean, indicating successful mean reversion, and the position is closed with a profit. Time-based exit occurs if the position remains open until the end of the month without achieving convergence, automatically closing the position regardless of profitability. This prevents positions from remaining open indefinitely when mean reversion fails to materialize.

Additionally, stop-loss thresholds can be implemented to limit downside risk when pairs diverge further instead of converging as expected. The stop-loss mechanism would monitor each individual pair's cumulative return on a daily basis. When any pair's performance drops below a predetermined threshold—typically set at 5%—the stop-loss is triggered and the position is closed to prevent further losses.

4.2 Portfolio Construction and Implementation

While the individual pair mechanics form the foundation of the strategy, practical implementation requires scaling this approach to the entire S&P 500 universe. At the beginning of each month, all possible stock pair combinations from the S&P 500 are evaluated, which theoretically could approach 124,750 potential pairs with 500 stocks.

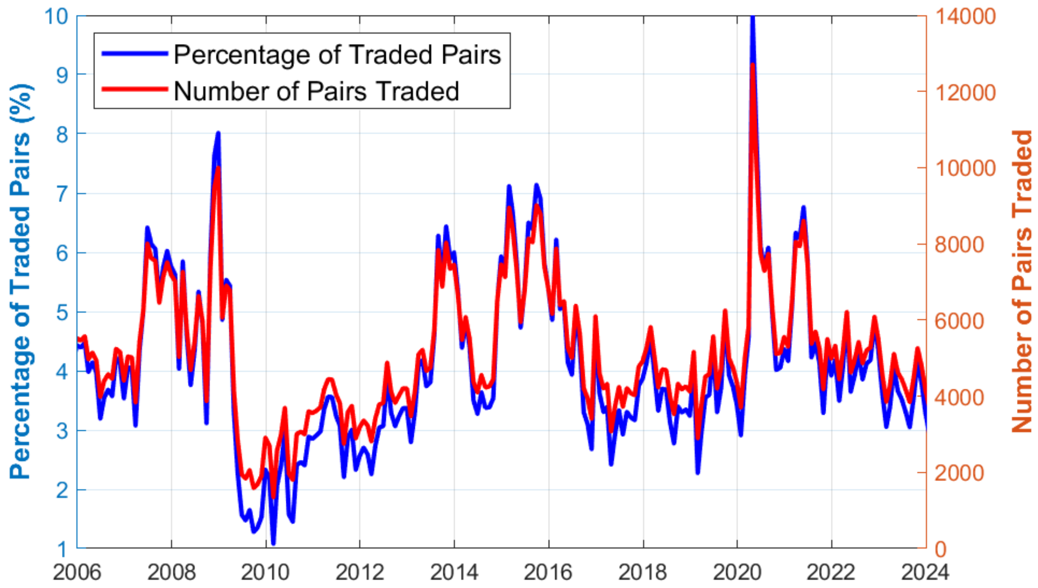


Figure 4: Monthly percentage of pairs traded and absolute number of pairs traded from the S&P 500 universe (2005-2024)

As illustrated in Figure 4, only a small percentage of these potential pairs meet the statistical criteria for trading in any given period. This selective approach ensures that only pairs with strong mean-reverting characteristics are included in the portfolio, though the number varies across different market environments.

For establishing a consistent baseline for analysis, the study defines a standard set of parameters that constitute the base case implementation:

Parameter	**Base Case Value**
Data Lookback Period	4 years
Maximum Holding Period	1 month
Entry Z-Score Threshold	± 2.0
Exit Z-Score Threshold	± 0.5
Stop-Loss Threshold	None
Weighting Scheme	Equal weighting

Table 1: Base case parameters for pairs trading strategy

These parameters serve as the control values against which variations are tested. The motivation for exploring parameter variations stems from several considerations. First, market efficiency and dynamics may change over time, necessitating different lookback periods to capture relevant historical relationships. Second, the sensitivity of entry and exit thresholds directly affects the trade-off between signal quality and opportunity frequency. Third, holding period constraints and stop-loss mechanisms balance the tension between allowing sufficient time for convergence and limiting exposure to adverse movements. By systematically varying these parameters as detailed in Table 2, the research aims to identify robust configurations that perform consistently across different market environments while addressing the specific research questions regarding temporal stability and strategy performance.

To identify optimal strategy parameters, the study employs a comprehensive grid search approach. The trading period is divided into a training period (2005-2014) and a testing period (2015-2024) to mitigate overfitting concerns.

The grid search explores 270 different parameter combinations across five key variables as detailed in Table 2:

Parameter	**Values**
Lookback Period Duration	2, 3, 4 years
Maximum Holding Period	1, 3, 4, 6, 12 months
Entry Z-Score Threshold	1, 2, 3
Exit Z-Score Threshold	± 0.1, ± 0.5
Stop-Loss Threshold	none , - 2.5%, - 5%

Table 2: Parameter variation tested in grid search optimization for pairs trading strategy (2005-2014)

Each combination is evaluated based on maximizing risk-adjusted performance through the Sharpe ratio. This optimization is conducted separately for different volume segments of the S&P 500 universe to address the research questions about how trading volume affects strategy performance.

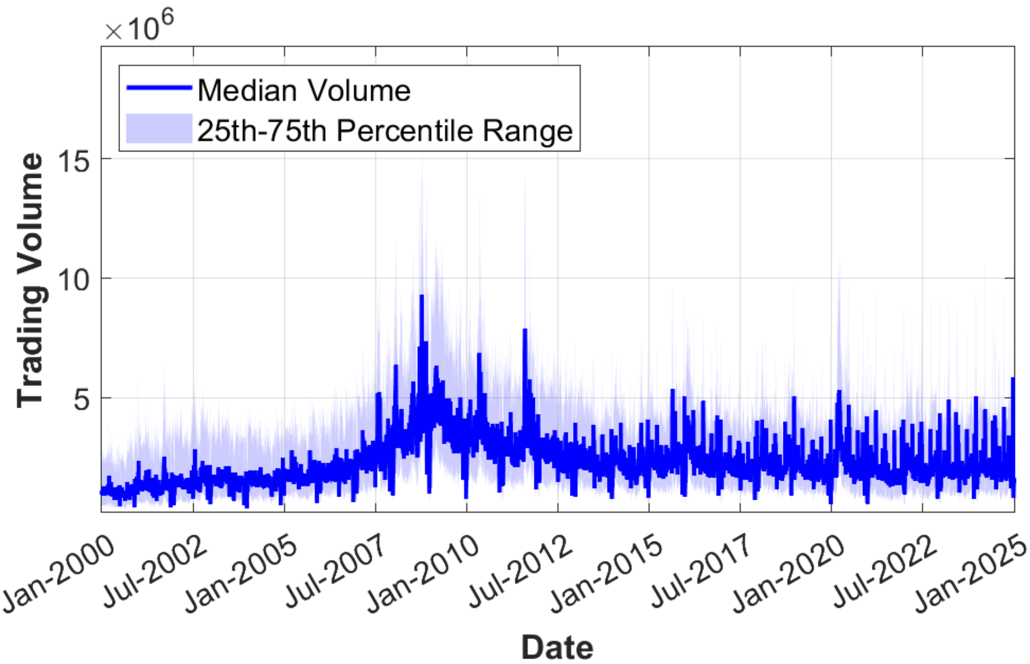


Figure 5: Daily trading volume distribution of S&P 500 stocks (2005-2024)

For position sizing and portfolio construction, several weighting schemes are implemented. Equal weighting assigns the same capital allocation to each pair trade, providing a baseline approach that does not favor any particular characteristics. Value weighting adjusts position sizes based on the market capitalization of the companies involved, allocating more capital to pairs involving larger companies which may have greater liquidity and stability. Volatility-based weighting allocates capital inversely proportional to the historical volatility of each pair, effectively equalizing the risk contribution of each position to the overall portfolio.

Additionally, an Ornstein-Uhlenbeck (OU) process-based weighting scheme is implemented. This approach models the spread as a mean-reverting process:

$$dZ_t = \theta(\mu - Z_t)dt + \sigma dW_t \tag{5}$$

where θ represents the speed of mean reversion. The half-life of mean reversion is calculated as $\ln(2)/\theta$, representing the expected time for the spread to revert halfway to its mean. Pairs with shorter half-lives receive higher weights, as they may offer more frequent profitable opportunities. The OU process is particularly suitable for pairs trading as it provides a mathematically tractable model for mean-reverting signals noted by Leung and Li (2015). As Tadi and Davallou (2021) demonstrated, using half-life values provides a valuable metric for asset selection in pairs trading strategies.

4.3 Performance Evaluation Framework

To thoroughly evaluate strategy performance across different conditions, the study employs several complementary metrics. Excess returns measures the strategy's return above the risk-free rate on both monthly and cumulative bases, providing insight into absolute performance over time. Sharpe ratio is calculated as the annualized average excess return divided by the standard deviation of returns:

$$\text{Sharpe ratio} = \frac{R_p - R_f}{\sigma_p} \tag{6}$$

where R_p represents the portfolio return, R_f is the risk-free rate, and σ_p is the standard deviation of portfolio returns. This ratio quantifies risk-adjusted performance, allowing for meaningful comparison across different strategy variants by standardizing returns per unit of risk. Maximum drawdown represents the worst peak-to-trough decline in portfolio value over the evaluation period:

$$\text{Maximum Drawdown} = \min\left[\left(\frac{V_t}{V_{\text{peak}}}\right) - 1\right] \tag{7}$$

Where V_t is the portfolio value at time t, and V_{peak} is the maximum portfolio value observed prior to time t. This metric captures downside risk that may not be fully reflected in standard deviation measures. Win rate is the percentage of periods that result in positive returns:

$$\text{Win Rate} = \frac{\text{Number of periods with positive returns}}{\text{Total number of periods}} \tag{8}$$

This indicates the consistency of the strategy's performance over time and helps assess whether returns are driven by frequent positive periods or less frequent but potentially larger gains concentrated in fewer periods.

5 Results

The findings from applying pairs trading to S&P 500 stocks from 2005 to 2024 reveal key insights about pairs trading performance. As market benchmarks, two distinct reference points provide context for the strategy's performance. The first benchmark is the market excess return (market return minus risk-free rate), representing broad market performance and serving as a passive investment alternative. The second benchmark is an equal-weighted long-short portfolio (Small minus Big, SMB) constructed from S&P 500 stocks, which provides a more relevant comparison as an active, market-neutral strategy. This SMB factor represents the performance difference between the smallest and largest stocks within the S&P 500 universe, adapting the original concept developed by Fama and French (1992, 1993) and later refined in their five-factor model (2015) to the specific stock universe. This benchmark is particularly appropriate for comparison because it shares several characteristics with pairs trading: it is a long-short strategy that is theoretically market-neutral, involves monthly rebalancing, and exploits a specific market anomaly (the size effect within large-cap stocks). Like pairs trading, this S&P 500-based SMB generates returns through relative value differences rather than absolute market direction, making it a natural benchmark for evaluating the effectiveness of statistical arbitrage strategies. Additionally, the size factor has been well-documented in academic literature as a persistent source of excess returns, see Banz (1981) and Reinganum (1981), providing an established baseline against which to measure pairs trading performance. These benchmarks serve as reference points throughout the analysis, allowing for comparisons with both the base case and optimized trading scenarios. For the base case (section 5.2), a performance benchmark is established using parameters commonly found in literature, though selected without extensive optimization—essentially representing a starting point rather than an optimized strategy. This base case is complemented by both CAPM and Fama-French five-factor model analyses to provide comprehensive risk-adjusted return assessments. Before proceeding with extensive parameter optimization, the 20-year dataset is split into distinct training and testing periods to evaluate strategy robustness. The base case performance across both periods establishes a reference for subsequent comparisons with optimized parameters. Thorough parameter optimization (section 5.3) follows, focused on maximizing Sharpe ratio in the training period, with evaluation of how these optimized parameters perform in the testing period. This approach helps distinguish between genuine strategy improvements and potential overfitting. Following parameter optimization, various weighting schemes (section 5.4) are evaluated as applied to the best Sharpe ratio parameter configuration across both periods, and transaction costs (section 5.5) are incorporated to provide realistic profitability assessments for practical implementation. These analyses collectively establish both the statistical and practical significance of the findings, with detailed examinations of each component following in their respective subsections.

5.1 Benchmark: SMB portfolio

Although SMB is well documented in the literature, it is presented here as a benchmark example for comparison purposes. Companies are sorted by market capitalization from smallest to largest, creating ten portfolios based on company size to track their performance. A clear pattern emerges across these portfolios: as company size increases, annualized returns consistently decreases. This observed inverse relationship between company size and returns aligns with the well-documented size effect in financial literature and validates the rationale for constructing a SMB benchmark, see the illustrative example in the Appendix in section 8.2.

The SMB long-short portfolio is constructed for the main trading period (2005-2024). Within each of the ten portfolios, stocks are equally weighted to avoid concentration in any particular company and to ensure that the size effect is captured purely through the number of securities rather than their individual market values. At the end of each month, all portfolios are recalculated and rebalanced based on the latest market capitalization rankings, ensuring the strategy continuously adapts to changing market conditions while maintaining equal weights within each decile.

The Small-Minus-Big (SMB) strategy within S&P 500 involves taking long positions in the smallest-cap companies (the smallest decile) while simultaneously shorting large-capitalization companies (the largest decile). This equal-weighted approach within each decile ensures that the portfolio effectively isolates the return premium associated with company size differences while maintaining market neutrality and avoiding the influence of individual stock performance on portfolio returns. This SMB portfolio serves as an active benchmark strategy for comparison with the pairs trading results, see Figure 6.

5.2 Baseline Strategy Performance

The initial analysis focuses on the unoptimized baseline approach: an equal-weighted portfolio with a 4-year lookback period, monthly rebalancing, and no stop-loss constraints.

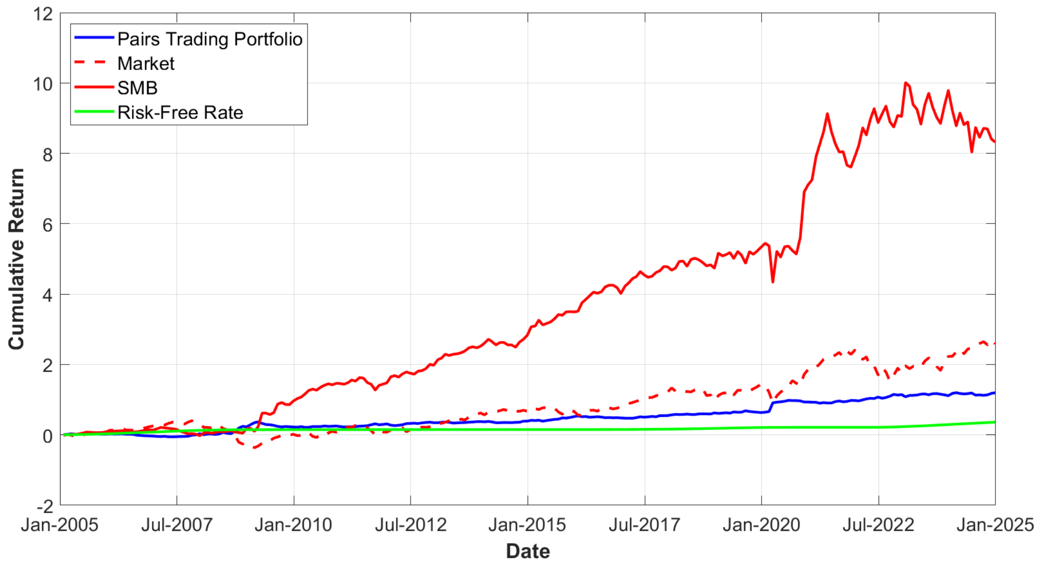


Figure 6: Cumulative returns comparison of base case pairs trading portfolio against market benchmarks (2005-2024).

Figure 6 displays the cumulative returns of the pairs trading strategy compared to the benchmarks over the 20-year study period (2005-2024). While the market and SMB portfolio achieve higher overall cumulative returns, the pairs trading strategy exhibits substantially lower volatility in its return pattern. This smoother performance trajectory highlights the strategy's potential as a alternative. The strategy shows exceptional performance during the 2008 financial crisis, where it not only maintained stability but actually generated positive returns while the market experienced significant drawdowns. Similarly, during the COVID-19 market disruption in 2020, the strategy again demonstrated its strength by producing substantial positive returns when the broader market suffered considerable losses. These crisis periods underscore the strategy's ability to capitalize on the increased market dislocations and volatility that typically accompany major market corrections, see section 5.2.3 for further details.

Metric	Base Portfolio	Market	SMB
Annualized Mean Return in %	4.11	7.65	12.22
Annualized Standard Deviation in %	5.93	15.68	14.63
Sharpe Ratio	0.69	0.49	0.84
Win Rate in %	58.75	62.08	60.00
Maximum Drawdown in %	11.27	54.67	19.54

Table 3: Comparison of portfolio and benchmark metrics (2005-2024)

Table 3 provides a comparison between the pairs trading portfolio, the market benchmark, and the SMB strategy. The performance metrics confirm the visual observations from the cumulative return chart. While the market delivers a higher annualized mean return of 7.65% compared to the portfolio's 4.11%, this comes at the cost of substantially higher risk. The pairs trading strategy exhibits only 38% of the market's volatility (5.93% vs. 15.68% annualized standard deviation), resulting in a superior Sharpe ratio of 0.69 versus the market's 0.49. Perhaps most striking is the maximum drawdown comparison—the pairs trading strategy experienced a maximum drawdown of just 11.27%, whereas the market suffered an 54.67% decline during severe market dislocations. The strategy demonstrates a modestly lower win rate of 58.75% compared to the market's 62.08%. The SMB benchmark shows the highest returns at 12.22% but with substantial volatility and a maximum drawdown of 19.54%, though it achieves the best Sharpe ratio of 0.84. These metrics collectively reinforce the pairs trading strategy's profile as a more consistent, lower-volatility approach with competitive risk-adjusted returns despite lower absolute performance.

5.2.1 CAPM Analysis and Implications

To assess how well the pairs trading strategy performs after accounting for market risk, the Capital Asset Pricing Model (CAPM) was applied. This involved a simple regression:

$$R_{\text{portfolio,t}} = \alpha + \beta \cdot R_{\text{market,t}} + \varepsilon_t \tag{9}$$

The scatterplot in Figure 7 of monthly returns shows a clear pattern where the strategy tends to move opposite to the market. The analysis found an annualized alpha of 5.28%, which is statistically significant at the 95% confidence level. This means the strategy generates substantial returns beyond what its market exposure would predict.

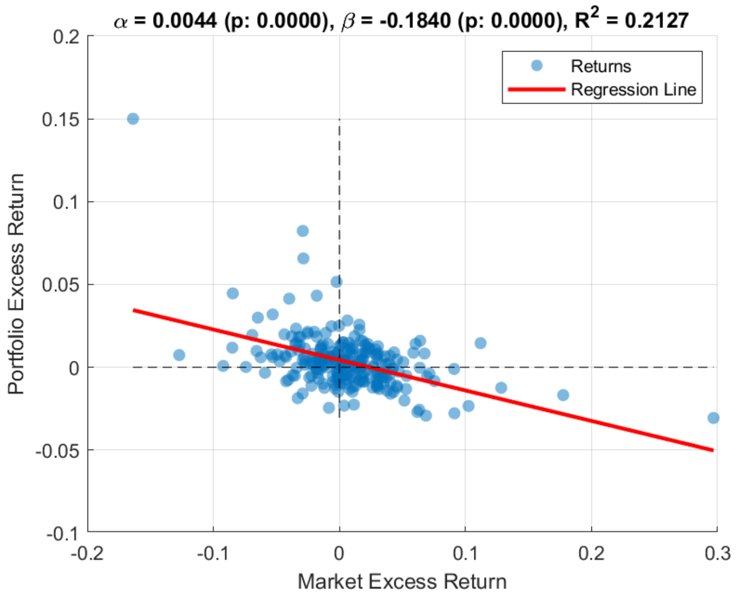


Figure 7: CAPM with monthly returns with an annualized alpha being 5.28%.

The beta coefficient is negative and statistically significant. This confirms that the strategy typically moves in the opposite direction of the broader market, especially during market downturns. This explains why the strategy performed so well during both the 2008 financial crisis and the 2020 COVID-19 market disruption.

The positive alpha combined with negative beta makes this strategy valuable for portfolio construction. The negative beta helps diversify traditional stock portfolios by reducing overall volatility more effectively than investments with zero market correlation. Since the strategy tends to perform well when markets fall, it serves as a partial hedge without the ongoing costs of typical hedging tools. Especially during market downturns, the portfolio demonstrates robust countercyclical performance, with exceptional instances of 15% returns during severe market declines (-15%), outperforming regression-based forecasts, as illustrated in Figure 7. The significant positive alpha shows the strategy finds pricing discrepancies that traditional risk factors can't explain. With a 5.28% annualized alpha, the strategy provides meaningful returns beyond its market exposure.

These results confirm what the cumulative return chart in Figure 6 shows and validate the pairs trading approach as both a source of returns and a portfolio diversifier. The significant alpha suggests that even in today's efficient markets, opportunities for statistical arbitrage still exist in large U.S. stocks, particularly during market disruptions.

5.2.2 Fama-French 5-factor model

To further investigate whether the strategy's performance can be attributed to common risk factors beyond market exposure, the Fama-French five-factor model (2015) is applied. This model extends CAPM by adding four additional factors: market risk premium (Mkt-RF), size premium (SMB), value premium (HML), profitability premium (RMW), and investment premium (CMA). This comprehensive framework helps determine whether the observed alpha genuinely represents market inefficiency exploitation or simply exposure to known risk factors.

The model is expressed as:

$$R_{\text{portfolio,t}} = \alpha + \beta_1 \cdot (Mkt_t - RF_t) + \beta_2 \cdot SMB_t + \beta_3 \cdot HML_t + \beta_4 \cdot RMW_t + \beta_5 \cdot CMA_t + \varepsilon_t \quad (10)$$

The regression results in Table 4 show the strategy maintains a statistically significant alpha of 0.41% monthly (approximately 5% annually) even after controlling for all five factors. The market factor (Mkt-RF) coefficient remains negative and highly significant (-0.15), confirming the countercyclical nature observed in the CAPM analysis.

Among the additional factors, HML shows a strong negative relationship (-0.26) while CMA demonstrates a positive relationship (0.17), both statistically significant at the 1% level. This suggests the strategy tends to perform better when growth stocks outperform value stocks and when companies with conservative investment policies outperform those with aggressive investment approaches. The SMB and RMW factors show no statistically significant relationship with the strategy's returns.

Factor	**Coefficient**	**Std Error**	**t-stat**	**p-value**
Alpha	0.0041	0.0009	4.6251	0.0000***
Mkt-RF	-0.1487	0.0212	-7.0129	0.0000***
SMB	-0.0118	0.0383	-0.3091	0.7576
HML	-0.2571	0.0360	-7.1486	0.0000***
RMW	0.0644	0.0498	1.2911	0.1979
CMA	0.1732	0.0576	3.0092	0.0029***
R-squared: 0.4213				

Table 4: Fama-French 5-factor regression results for pairs trading portfolio

The residuals analysis in Figure 8 reveals that during extreme market downturns, the strategy produces unexplained positive returns beyond what the factor models predict. This critical pattern appears in the left portion of the plot, where market returns are strongly negative but residuals are positive and often substantial. This asymmetric response to market conditions further confirms the strategy's value as a portfolio hedge during market stress periods.

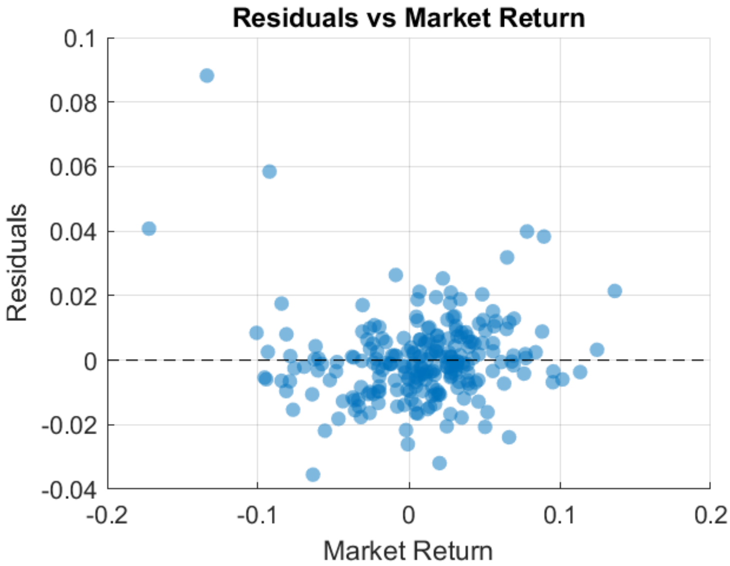


Figure 8: Comparison of residuals and market return showing higher positive residuals during negative market returns

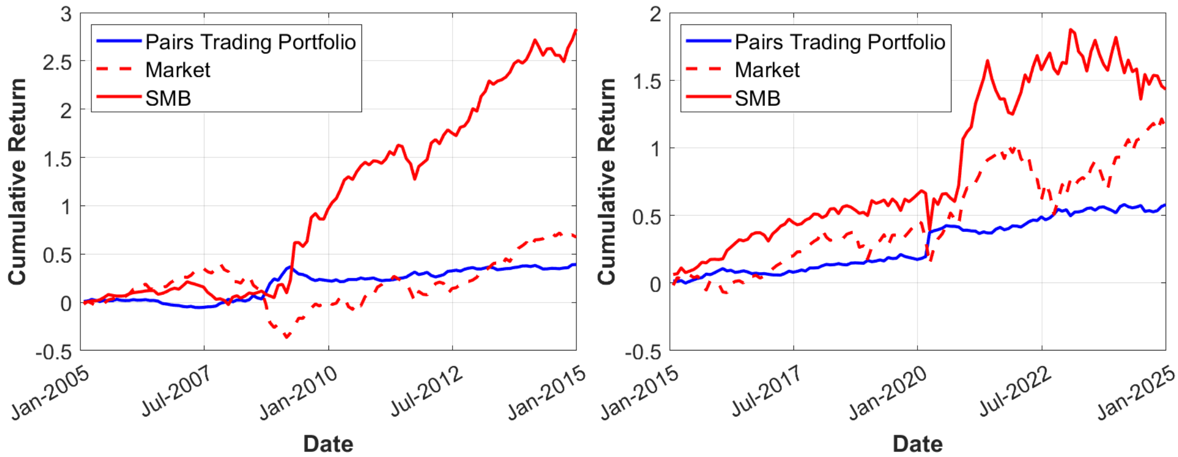
Overall, the persistence of significant alpha in both CAPM and FF5 models indicates that pairs trading captures market inefficiencies not explained by common risk factors, demonstrating its potential as both a return enhancer and diversification tool in portfolio construction.

5.2.3 Base Case in Two Periods

To evaluate the robustness of the pairs trading strategy, the 20-year dataset is divided into two equal periods: Period 1 (2005-2014) and Period 2 (2015-2024). This division enables assessment across different market regimes, establishes a framework for parameter optimization with training and testing periods, and examines whether market efficiency has increased over time.

In Period 1, the SMB benchmark significantly outperformed the pairs trading strategy (14.48% vs. 3.47% annual return), but with substantially higher volatility (14.69% vs. 5.87%). The pairs trading strategy maintained remarkable stability during the 2008-2009 financial crisis while capturing less of the subsequent rally, see Figure 9a. The Market benchmark delivered moderate returns of 6.52% with high volatility of 16.31%.

Period 2 presents a similar dynamic. Both benchmarks outperform the pairs trading portfolio in terms of cumulative returns. The pairs trading strategy delivered steady returns (4.74% annually) with a Sharpe ratio of 0.79, while the benchmarks showed mixed performance. The SMB portfolio's returns declined to 9.96% with a lower Sharpe ratio of 0.68, while the Market benchmark's returns increased to 8.78% with an improved Sharpe ratio of 0.58. The strategy's maximum drawdown remained exceptionally low at 4.34%, compared to the Market's 25.89% and SMB's 17.93% declines.



(a) 2005-2014: SMB outperforms with higher volatility. (b) 2015-2024: Pairs Trading steady without major drops.

Figure 9: Pairs trading and SMB market returns split in two periods.

Table 5 highlights several critical insights about the strategy's temporal evolution. First, the pairs trading approach demonstrates remarkable consistency in its risk profile across both periods, with only a slight increase in volatility (5.87% to 6.01%). Both benchmarks maintain relatively stable volatility levels across periods. Second, the strategy's risk-adjusted performance improves substantially in Period 2, with its Sharpe ratio increasing from 0.59 to 0.79, while the SMB benchmark's Sharpe ratio decreases from 0.99 to 0.68, and the Market benchmark's ratio increases from 0.40 to 0.58. Third, the win rate metrics reveal the strategy's consistency, maintaining stable winning percentages across both periods (58.33% to 59.17%), while the Market benchmark's win rate improves from 59.17% to 65.00%, and SMB's win rate declines from 62.50% to 57.50%. Perhaps most striking is the maximum drawdown comparison, where the pairs trading strategy's already conservative risk profile dramatically improves in Period 2 (11.27% to 4.34%), while both benchmarks experience substantial maximum losses in both periods, with the Market showing 54.67% to 25.89% and SMB showing 19.53% to 17.93%.

Metric	Period 1 (2005-2014)			Period 2 (2015-2024)		
	Portfolio	Market	SMB	Portfolio	Market	SMB
Annualized Mean Return in %	3.47	6.52	14.48	4.74	8.78	9.96
Annualized Standard Deviation in %	5.87	16.31	14.69	6.01	15.08	14.61
Sharpe Ratio	0.59	0.40	0.99	0.79	0.58	0.68
Win Rate in %	58.33	59.17	62.50	59.17	65.00	57.50
Maximum Drawdown in %	11.27	54.67	19.53	4.34	25.89	17.93

Table 5: Performance metrics comparison of base case pairs trading portfolio versus benchmark portfolios across training and testing periods.

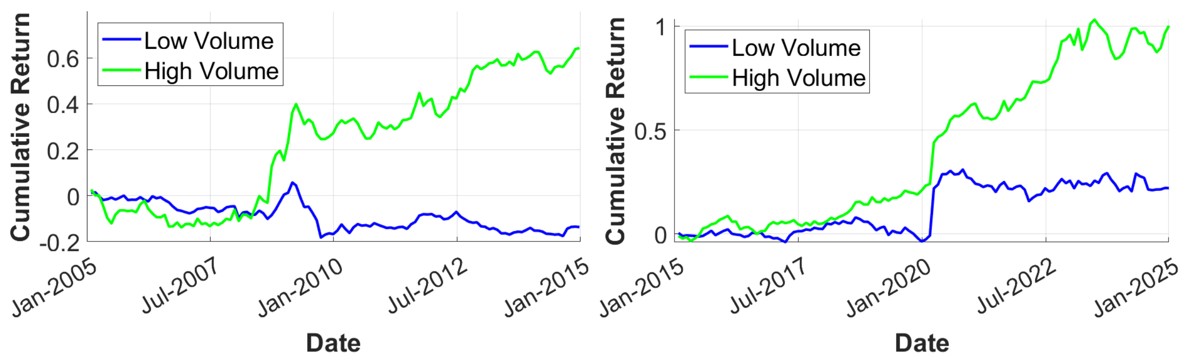
This analysis demonstrates the strategy's consistency and improving risk-adjusted performance across varying market conditions. While the SMB benchmark delivered higher absolute returns in both periods, it came with significantly higher risk and drawdowns. The Market benchmark showed moderate performance with high volatility throughout both periods. The pairs trading strategy's consistent low-volatility profile and dramatically improved drawdown characteristics in Period 2 indicate genuine strategy merit focused on capital preservation rather than data mining or overfitting—establishing a solid foundation for subsequent parameter optimization.

5.2.4 Volume Effects on the Base Case

To understand how liquidity impacts pairs trading performance, the base case strategy is analyzed across different volume environments. Rather than applying the strategy to the entire market, the universe is segmented into volume deciles and performance is examined in the lowest and highest volume deciles while maintaining identical parameter settings from the base case. Each decile contains approximately 50 stocks, creating 1,225 potential pairs (roughly 1% of the total pair pool). This approach isolates the pure effect of trading volume on strategy effectiveness, controlling for all other factors.

The volume analysis reveals striking patterns in pairs trading performance across different liquidity environments. The results in Figure 10 demonstrate a clear volume premium, where high-volume pairs consistently outperform both the overall portfolio and low-volume pairs across both time periods.

In Period 1 (2005-2014), the performance disparity is dramatic. High-volume pairs generated strong returns of 5.48% annually, significantly outperforming the overall portfolio (3.47%) and dramatically exceeding low-volume pairs, which actually posted negative returns of -1.27%. This 6.75 percentage point spread between high and low volume pairs highlights the critical importance of liquidity in pairs trading. The low-volume pairs also exhibited higher volatility (6.29%) relative to their negative returns, resulting in a negative Sharpe ratio of -0.20, while high-volume pairs maintained reasonable risk-adjusted performance with a Sharpe ratio of 0.53, see Table 6.



(a) Training period (2005-2014)

(b) Testing period (2015-2024)

Figure 10: Cumulative returns comparison between low-volume and high-volume pairs trading portfolios with base case parameter setup for (a) training period 2005-2014 and (b) testing period 2015-2024.

Period 2 (2015-2024) shows an even more pronounced volume effect. High-volume pairs delivered exceptional performance with 7.25% annual returns and an outstanding Sharpe ratio of 0.93, while low-volume pairs managed only 2.40% returns with a modest Sharpe ratio of 0.26. The cumulative return differential over the decade is substantial, with high-volume pairs achieving 100% cumulative returns compared to just 22% for low-volume pairs. Notably, high-volume pairs also demonstrated superior risk management, with lower volatility (7.81% vs 9.34%) and higher win rates (66.67% vs 50.00%) compared to their low-volume counterparts.

Metric	Period 1 (2005-2014)			Period 2 (2015-2024)		
	All Vol.	Low Vol.	High Vol.	All Vol.	Low Vol.	High Vol.
Annualized Mean Return in %	3.47	-1.27	5.48	4.74	2.40	7.25
Annualized Std Dev. in %	5.87	6.29	10.25	6.01	9.34	7.81
Sharpe Ratio	0.59	-0.20	0.53	0.79	0.26	0.93
Win Rate in %	58.33	50.00	59.17	59.17	50.00	66.67
Maximum Drawdown in %	11.27	22.66	16.06	4.34	11.64	9.36
Cumulative Returns in %	39	-14	64	58	22	100

Table 6: Performance metrics comparison of base case portfolio for the whole universe of S&P 500 pairs versus low-volume and high-volume pairs trading portfolios across training and testing periods

Given these substantial performance differences across volume environments, the next step involves systematically varying input parameters for both low and high volume deciles to optimize strategy performance within each liquidity regime.

5.3 Parameter variation

Following the baseline analysis, a comprehensive parameter optimization was conducted using a rigorous out-of-sample approach: the best parameter setup was identified using data from the first period (2005-2014), and this optimal configuration was then tested on the second period (2015-2024) to validate performance consistency and avoid overfitting bias. The grid search explored 270 distinct parameter combinations across lookback periods, holding periods, entry/exit thresholds, and stop-loss settings, performed separately for low-volume stocks (bottom decile) and high-volume stocks (top decile) to test volume-related effects, see Table 2.

5.3.1 Sharpe Ratio Optimization

Optimizing for risk-adjusted returns reveals patterns similar to return maximization (see in Appendix 8.3) but with important distinctions. Figure 11 shows Sharpe ratio distributions across parameter combinations for both volume segments. Low-volume pairs cluster tightly around -0.25, while high-volume pairs show a broader distribution centered around 0.2, with top combinations reaching beyond 2. This strong separation between the distributions provides strong evidence that volume is a fundamental driver of strategy effectiveness. This disparity in risk-adjusted performance directly addresses the research question regarding the relationship between trading volume and pairs trading profitability. The consistency across both raw return and risk-adjusted metrics confirms that volume significantly influences strategy effectiveness, with high-volume stocks offering superior opportunities for pairs trading strategies.

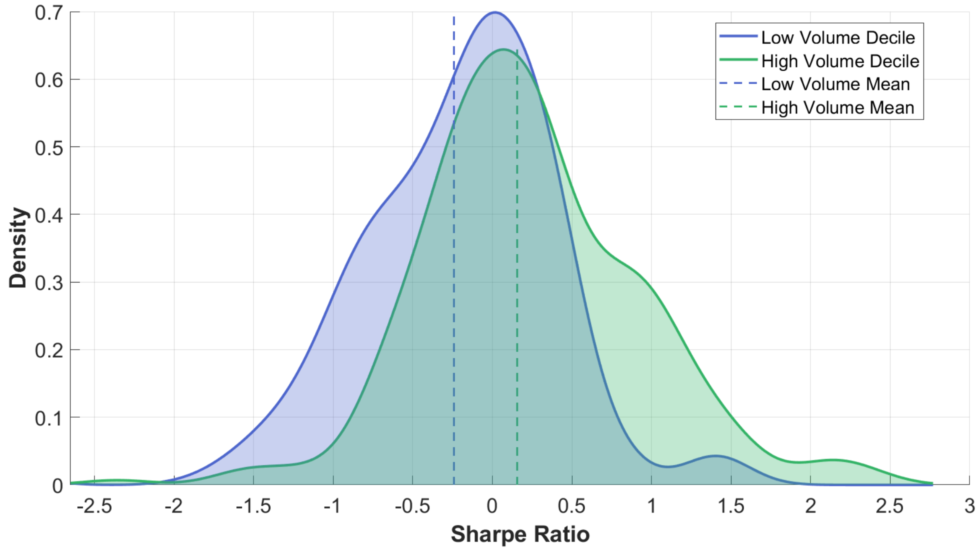
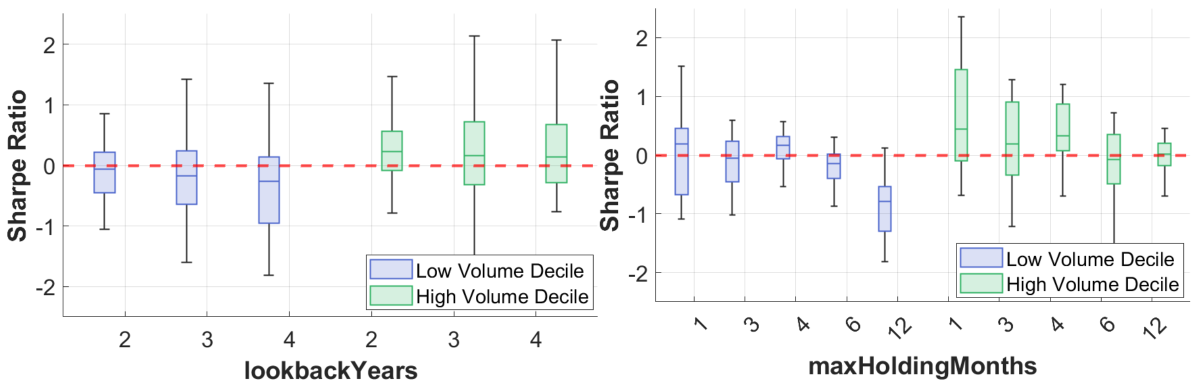


Figure 11: Sharpe ratio distributions for low-volume versus high-volume pairs trading portfolios while performing parameter variation (2005-2014).

Figure 12 presents box plots comparing Sharpe ratios across different lookback periods (panel a) and maximum holding periods (panel b) for both high-volume (green) and low-volume (blue) stock pairs. The figure clearly illustrates the consistent performance gap between volume segments across all parameter settings.



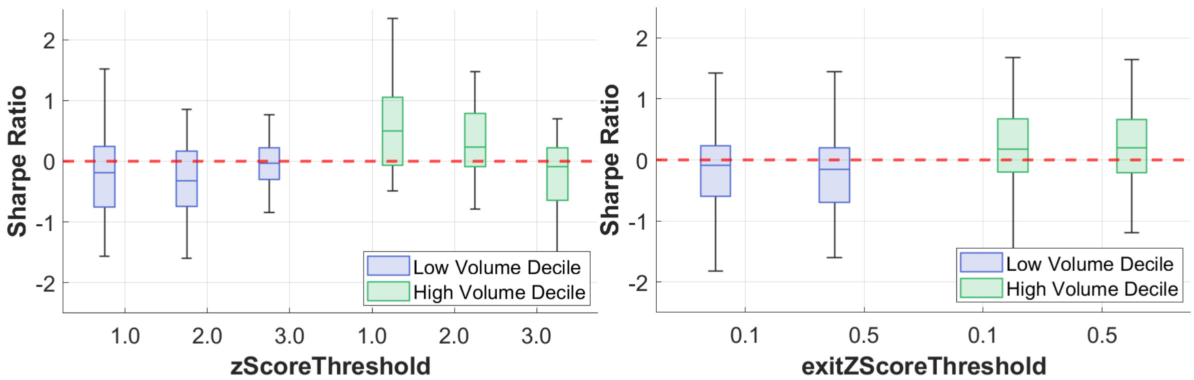
(a) Lookback years of data

(b) Maximum holding period of pairs in months

Figure 12: Sharpe ratio distributions across parameter variations for low-volume and high-volume pairs trading portfolios (2005-2014)

The figure reveals that for high-volume pairs, longer holding periods systematically reduce Sharpe ratios, advocating for shorter holding periods that capture quick mean reversions before redeploying capital. The lookback period shows minimal impact on performance, with high-volume pairs consistently producing higher Sharpe ratios regardless of the specific lookback setting. Across all parameters, high-volume pairs consistently deliver superior Sharpe ratios compared to low-volume pairs, which rarely achieve positive values regardless of settings.

Figure 13 displays box plots comparing Sharpe ratios across entry z-score thresholds (panel a) and exit z-score thresholds (panel b), again distinguishing between high-volume (green) and low-volume (blue) stock pairs. The figure shows the persistent performance advantage of high-volume pairs across all threshold settings, with high-volume pairs achieving notably higher Sharpe ratios while low-volume pairs consistently remain near or below zero.



(a) Entry z-score (b) Exit z-score

Figure 13: Sharpe ratio distributions across parameter variations for low-volume and high-volume pairs trading portfolios (2005-2014)

For high-volume pairs, all exit z-score threshold settings produce consistently high Sharpe ratios with minimal differences between them, indicating that the exit threshold is less critical than other parameters as long as high-volume stocks are used. For entry z-score thresholds, while high-volume pairs generally outperform low-volume pairs across all settings, the lower entry z-score threshold (1) yields marginally better results among high-volume pairs compared to higher thresholds (2 and 3). This suggests that entering positions earlier when smaller divergences occur provides a small additional advantage by capturing more trading opportunities, while waiting for extremely high divergences may indicate structural breaks or permanent shifts rather than temporary mean-reverting opportunities.

Figure 14 displays box plots comparing Sharpe ratios across different stop-loss threshold settings (no stoploss, - 5% and - 2.5%) for both high-volume (green) and low-volume (blue) stock pairs. The figure demonstrates the familiar pattern of high-volume pairs substantially outperforming low-volume pairs across all stop-loss settings, with high-volume pairs with stoploss achieving Sharpe ratios consistently above 0.5 while low-volume pairs remain clustered around zero. Notably, the tighter the stop-loss threshold, the better the Sharpe ratio for both volume categories, with the most restrictive 2.5% stop-loss delivering the highest performance.

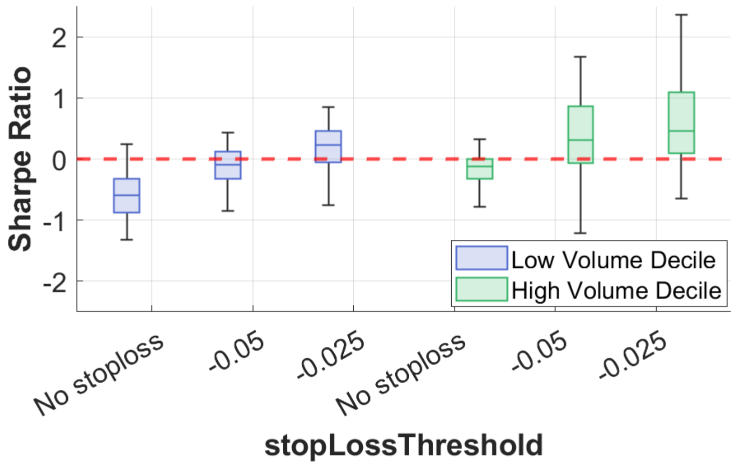
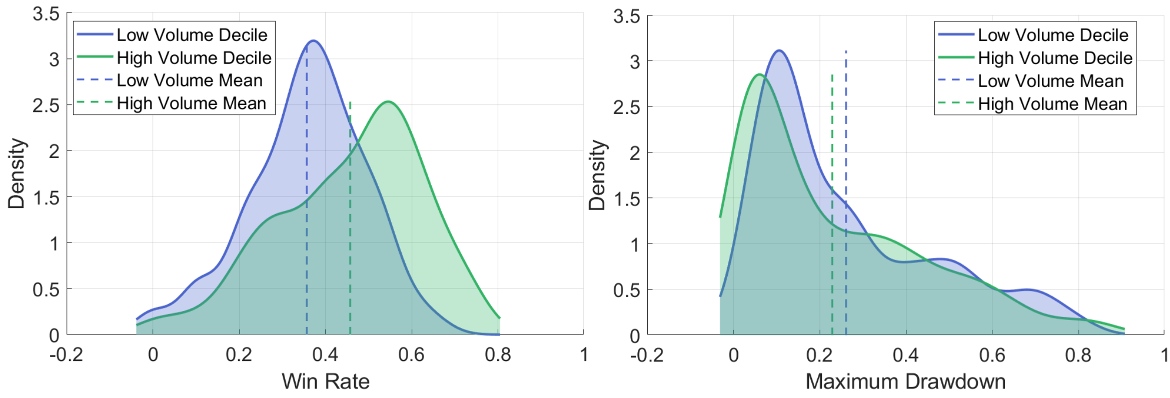


Figure 14: Sharpe ratio distributions across stop-loss threshold variations for low-volume and high-volume pairs trading portfolios (2005-2014)

The analysis reveals that stop-loss thresholds have an enhancing impact on performance for both volume segments. For high-volume pairs, all stop-loss settings produce similarly strong results, with a clear preference for tighter stop-loss thresholds, as the -2.5% setting yields notably higher median Sharpe ratios compared to the -5% threshold scenario. For low-volume pairs, the pattern is even more pronounced, with performance improving progressively as stop-loss thresholds become more restrictive. This demonstrates that tighter stop-loss mechanisms provide effective downside protection across both volume categories. The parameter optimization results suggest that the ideal pairs trading strategy would focus on high-volume stocks, employ shorter holding periods, and utilize smaller stop-loss boundaries.

Further analysis of key performance metrics reveals significant differences between high and low volume deciles. Figure 15 compares the distribution of win rates, highlighting the superior consistency of high-volume pairs. The high-volume decile exhibits a mean win rate of 45% across parameter combinations, with optimal configurations achieving exceptional win rates exceeding 80%. In contrast, the low-volume decile shows a substantially lower mean win rate of 35%, with even the best combinations rarely exceeding 70%. This win rate differential underscores the greater predictability of mean reversion in high-volume pairs. The notably higher success rate suggests that price relationships between liquid stocks tend to be more reliable and less subject to extended dislocations that prevent convergence within the strategy's timeframe.

Maximum drawdown distributions further demonstrate the advantageous risk profile of high-volume pairs. While the distributions are fairly similar in shape, the high-volume decile exhibits a notably lower mean maximum drawdown compared to low-volume pairs. Both distributions show comparable behavior in their right tails, but the key advantage lies in high-volume pairs' consistently lower average drawdowns, indicating better overall risk control across parameter settings.



(a) Win rate by volume decile

(b) Maximum drawdown by volume decile

Figure 15: Distribution of trading performance metrics comparing low-volume and high-volume pairs trading portfolios (2005-2014)

These metrics collectively reinforce the finding that high-volume pairs not only offer superior return potential and risk-adjusted performance but also provide greater consistency in trade outcomes with generally more favorable drawdown characteristics when properly parameterized.

5.3.2 Lower Volume Decile

Table 7 presents the optimal parameters for the lower volume decile when maximizing Sharpe ratio. This configuration features a medium lookback period (2 years), short holding period (1 month), low entry threshold (z-score of 1), tight exit threshold (0.1), and tighter stop-loss (-2.5%).

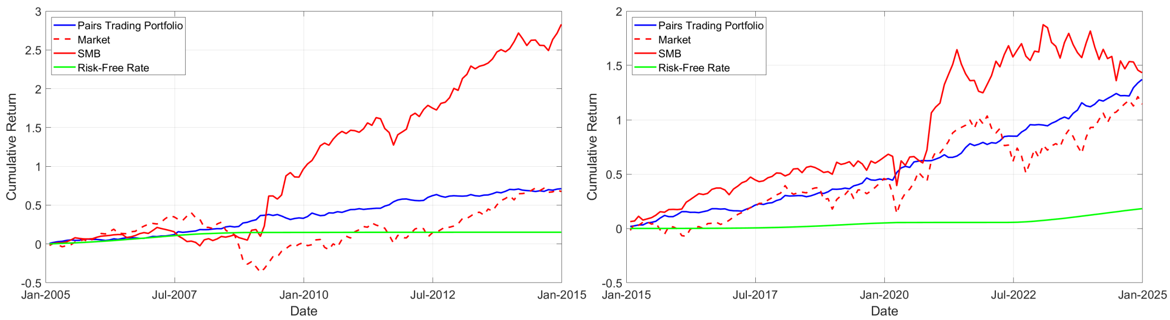
Parameter/Metric	**Training (2005-2014)**	**Testing (2015-2024)**
Lookback Years	2	2
Max Holding Months	1	1
Entry Z-score	1	1
Exit Z-score	0.1	0.1
Stop-loss in %	- 2.5	- 2.5
Annualized Return in %	5.5	8.8
Annualized STD in %	3.60	4.1
Sharpe Ratio	1.51	2.15
Win Rate in %	66.7	70.0
Max Drawdown in %	4.7	1.8
Cumulative Return in %	71.2	137.1
Mean (number of pair trades/month)	269	239
Std dev. (number of pair trades/month)	69.1	65.3

Table 7: Optimal parameter configuration and performance metrics for low-volume pairs trading strategy across training and testing periods.

Remarkably, performance improved in the testing period compared to the training period. The annualized return increased substantially from 5.5% to 8.8%, while the Sharpe ratio increased from 1.51 to 2.15. The win rate rose from 66.7% to 70%, and maximum drawdown decreased from 4.7% to just 1.8%.

These results contrast sharply with the return-maximizing approach, suggesting that optimizing for risk-adjusted performance produces more robust strategies that can adapt to changing market conditions, see 8.3. The improved testing period performance indicates that Sharpe ratio optimization captured fundamental relationships rather than overfitting to historical patterns. Figure 16 visually confirms this improvement, with the pairs trading portfolio (blue line) demonstrating steadier growth in the testing period (panel b) compared to the training period (panel a). While the market benchmark and SMB portfolio (red lines) outperform the pairs trading portfolio in terms of cumulative returns during the training period (panel a), the pairs trading strategy demonstrates more competitive performance in the testing period (panel b). Throughout both periods, the benchmark portfolios achieve higher returns at certain intervals but also experience significant volatility, highlighting the fundamental trade-off between higher potential returns and increased risk exposure compared to the pairs trading strategy's consistent, lower-volatility growth pattern.

While the earlier analysis demonstrated that high-volume stocks generally provide superior pairs trading opportunities across most parameter combinations, these optimized results for low-volume securities reveal an important nuance. With carefully calibrated parameters specifically tuned to their characteristics, low-volume stocks can also deliver compelling risk-adjusted returns. The optimal configuration for low-volume pairs turns out to be identical to that of high-volume pairs, with both requiring the same exit thresholds and stop-loss parameters. This alignment suggests that the pairs trading strategy maintains consistent parameter requirements across different liquidity environments, with execution risk and liquidity constraints having less impact on optimal parameter selection than initially anticipated.



(a) Training period (2005-2014)

(b) Testing period (2015-2024)

Figure 16: Cumulative return of low volume decile pairs during training and testing period.

The cumulative return nearly doubled from 71.2% in the training period to an impressive 137.1% in the testing period, demonstrating the strategy's ability to adapt to changing market dynamics. These findings strengthen the case for Sharpe ratio optimization as the preferred approach for parameter selection, as it appears to identify more fundamentally sound relationships between pairs rather than capturing temporary anomalies (see section 8.3).

5.3.3 Upper Volume Decile

Now optimization approach is examined when applied specifically to high-volume securities, which the earlier analysis identified as offering superior pairs trading opportunities (see section 5.3.1).

Table 8 presents the optimal parameters for the upper volume decile when maximizing Sharpe ratio. This configuration features the lookback period (2 years), shortest holding period (1 month), lowest entry threshold (z-score of 1), exit threshold (0.1), and tight stop-loss (-2.5%). Similar to the lower volume decile, performance improved during the testing period. Annualized return increased from 14.43% to 18.52%, while the Sharpe ratio rose from an already impressive 2.35 to 2.72. The win rate showed remarkable improvement from 75% to 80.8%, indicating exceptional consistency in identifying profitable opportunities. Maximum drawdown decreased from 1.65% to 1.24%, and the overall risk profile, which was already very good, improved further.

Parameter/Metric	**Training (2005-2014)**	**Testing (2015-2024)**
Lookback Years	2	2
Max Holding Months	1	1
Entry Z-score	1	1
Exit Z-score	0.1	0.1
Stop-loss in %	- 2.5	- 2.5
Annualized Return in %	14.43	18.52
Annualized STD in %	6.15	6.79
Sharpe Ratio	2.35	2.73
Win Rate in %	75	80.8
Max Drawdown in %	1.65	1.24
Cumulative Return in %	312	515
Mean (number of pair trades/month)	270	260
Std dev. (number of pair trades/month)	57.3	54.1

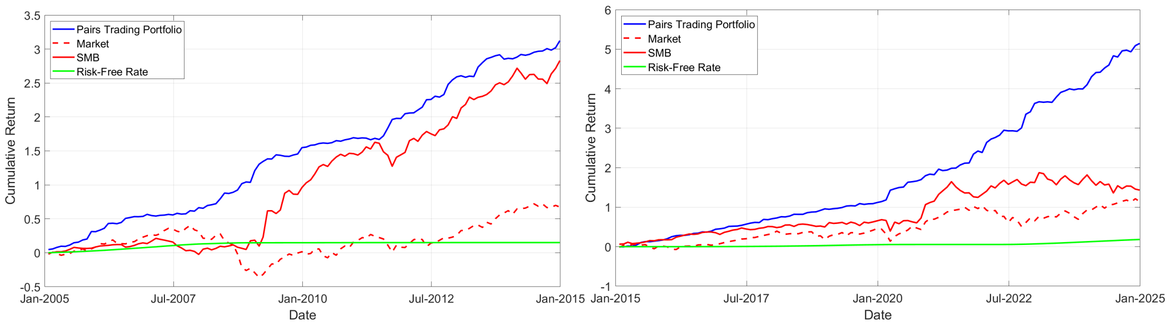
Table 8: Optimal parameter configuration and performance metrics for high-volume pairs trading strategy across training and testing periods.

Figure 17 visually confirms the superior performance of the high-volume pairs strategy, particularly in the testing period (panel b). In the training period (panel a), the pairs trading portfolio (blue line) consistently outperformed both the market and SMB benchmarks throughout most of the period, achieving steady growth that culminated in approximately 312% cumulative returns by 2014. The testing period reveals an even more impressive performance. While both the market and SMB benchmarks experienced significant volatility and modest returns, with the SMB benchmark showing particular weakness after

2021, the optimized high-volume pairs trading strategy demonstrated remarkably consistent, almost linear growth. The pairs trading portfolio continued its upward trajectory throughout the testing period, ultimately achieving cumulative returns of approximately 515% and substantially outperforming both the SMB and market benchmarks by the end of the period.

These results significantly outperform both the base case and the optimized low-volume strategy, confirming the research question regarding whether high-volume securities offer inherently better pairs trading opportunities. The consistent performance across both training and testing periods is particularly noteworthy, as it suggests the strategy is capturing fundamental market relationships rather than transient patterns or data artifacts.

When comparing the optimal parameters between high-volume and low-volume segments, the parameters are identical across both volume categories: a 2-year lookback period, 1-month holding period, and a tight -2.5% stop-loss threshold, demonstrating the robustness of the pairs trading strategy across different liquidity environments. This parameter combination reveals a strategy designed to capture quick market changes effectively, where the 2-year lookback period ensures the analysis relies on fresh, recent market data that reflects current relationships between stock pairs, the short 1-month holding period allows for rapid position turnover to capitalize on temporary price divergences, and the tight -2.5% stop-loss threshold provides immediate protection against trades that deviate significantly from expected patterns. This configuration creates a nimble trading approach that adapts quickly to market dynamics while maintaining strict risk control, explaining its effectiveness across both volume segments.



(a) Training period (2005-2014)

(b) Testing period (2015-2024)

Figure 17: Cumulative return of high volume decile during training and testing period.

The exceptional win rate of 80% in the testing period for high-volume pairs demonstrates that pairs trading can achieve remarkable consistency even in changing market environments. This high success rate, combined with controlled drawdowns and strong risk-adjusted returns, confirms that Sharpe ratio optimization produces robust strategies that maintain effectiveness beyond their training period.

5.4 Weighting methods

For the subsequent analyses for the weighting schemes and transaction costs, the optimal parameter configuration from high-volume stocks will be implemented, which delivered the highest Sharpe ratio (2.73) and consistent performance metrics in the testing period.

Different weighting schemes have been implemented to assess their impact on strategy performance. Equal weighting assigns identical weight to each pair regardless of characteristics, while value weighting allocates capital proportionally to the combined market capitalization of both stocks. Risk parity weighting gives higher allocations to less volatile pairs to equalize risk contribution across the portfolio. Half-life weighting favors pairs with faster mean reversion by assigning greater weight to shorter half-lives, and volume weighting increases allocation to pairs with higher average trading volumes.

Figure 18 compares the cumulative returns of these weighting strategies during the training period (2005-2014) for high-volume stocks. All weighting methods demonstrate similar pattern dynamics, showing strong performance during the 2008-2009 financial crisis when the market (dashed line) experienced significant drawdowns. Following the crisis, all weighting schemes delivered steady returns with minimal drawdowns through 2015.

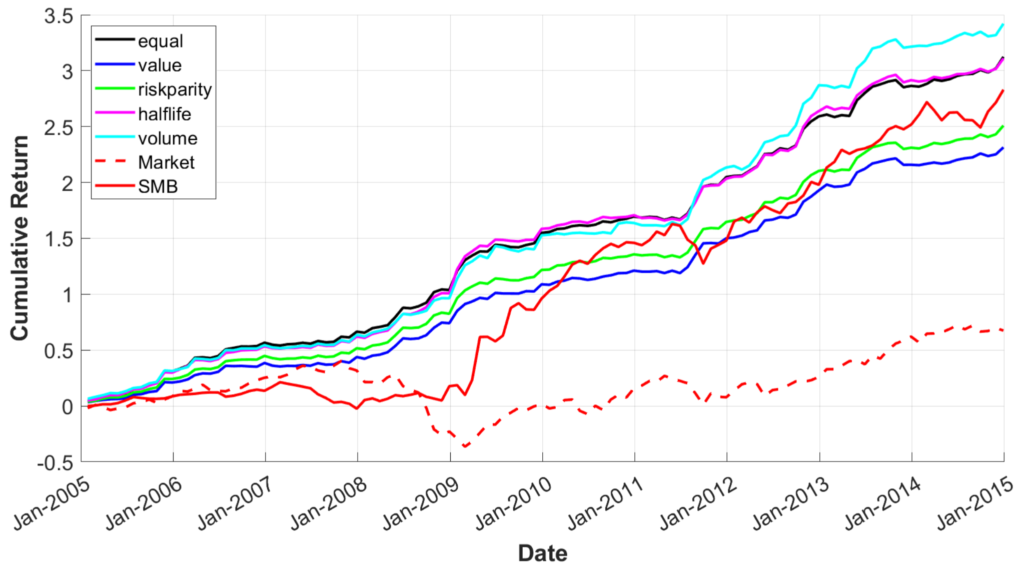


Figure 18: Cumulative returns comparison of different weighting methodologies for high-volume pairs trading portfolios during the training period (2005-2014).

The performance metrics in Table 9 reveal subtle differences among the approaches. Volume weighting produced the highest annualized return (15.20%) but with correspondingly higher volatility (7.14%), while risk parity delivered the highest Sharpe ratio (2.36) alongside the lowest volatility (5.42%), demonstrating its effectiveness in optimizing risk-adjusted returns. Equal weighting achieved a strong balance with the second-highest Sharpe ratio (2.35) and competitive returns (14.43%). Both risk parity and half-life weighting achieved the highest win rates (75.83%), while equal weighting maintained the lowest maximum drawdown (1.65%). Value weighting delivered solid performance across all metrics, maintaining a strong

Sharpe ratio of 2.23 despite being the most conservative approach. While all weighting schemes delivered exceptionally strong risk-adjusted performance with Sharpe ratios exceeding 2.0, risk parity marginally outperformed the others in terms of risk-adjusted returns, followed closely by equal weighting. This suggests that the primary driver of performance is the selection of high-volume pairs with optimized parameters, with weighting methodology playing a secondary role.

Metric	**Equal**	**Value**	**Risk Parity**	**Halflife**	**Volume**
Annualized Mean Return in %	14.43	12.19	12.76	14.41	15.20
Annualized Standard Deviation in %	6.15	5.46	5.42	6.42	7.14
Sharpe Ratio	2.35	2.23	2.36	2.25	2.13
Win Rate in %	75.00	72.50	75.83	75.83	69.17
Maximum Drawdown in %	1.65	2.28	2.07	1.74	1.96

Table 9: Performance metrics comparison across different weighting methodologies for high-volume pairs trading portfolios during the training period (2005-2014).

Figure 19 compares the cumulative returns of these weighting strategies during the testing period (2015-2024) for high-volume stocks. All weighting methods demonstrate consistent upward trends, with particularly strong performance during the COVID-19 market disruption in 2020. The pairs trading portfolios consistently outperformed the market throughout this volatile period, while the market benchmark (dashed line) experienced significant volatility. Meanwhile, the SMB benchmark stagnated after the COVID-19 pandemic, showing little growth in the post-2020 period, further highlighting the robust performance of the pairs trading strategies across different market conditions.

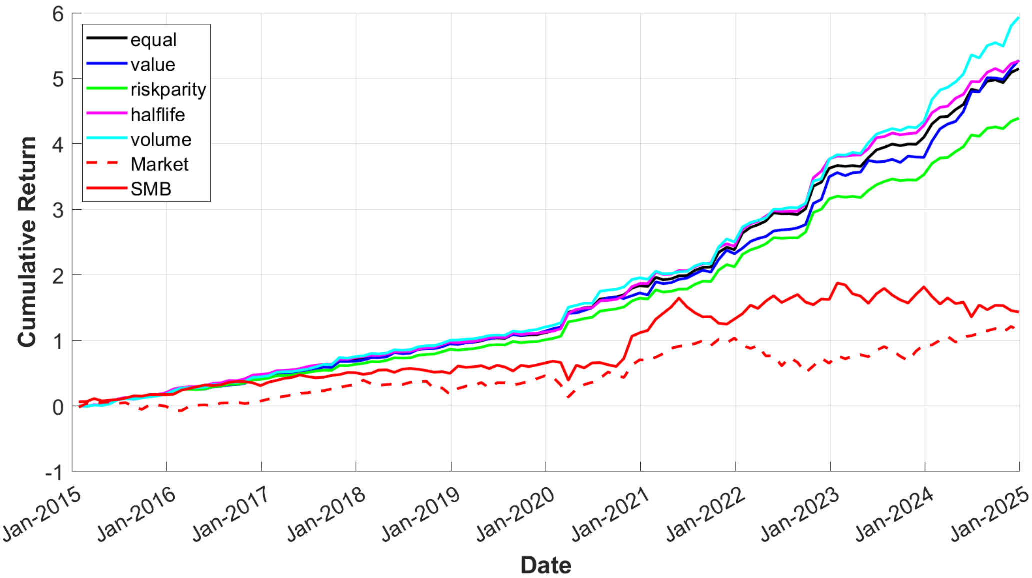


Figure 19: Cumulative returns comparison of different weighting methodologies for high-volume pairs trading portfolios during the testing period (2015-2024).

The performance metrics in Table 10 for the testing period reveal that volume weighting achieved the highest annualized return (19.78%) while risk parity delivered the highest Sharpe ratio (2.74), closely followed by equal weighting (2.73). Half-life weighting demonstrated the strongest consistency with the highest win rate (85.83%) and lowest maximum drawdown (1.19%). Equal weighting maintained excellent performance with strong returns (18.52%) and the second-highest Sharpe ratio, while also achieving a high win rate (80.83%) and low maximum drawdown (1.24%). While all weighting schemes delivered exceptionally strong performance with Sharpe ratios exceeding 2.6, risk parity marginally outperformed the others in terms of risk-adjusted returns, with equal weighting performing nearly as well. The close performance across all methodologies, with differences in Sharpe ratios within 0.1, demonstrates that the primary driver of performance remains the selection of high-volume pairs with optimized parameters, with weighting methodology providing only marginal differentiation in this high-performance environment.

Metric	**Equal**	**Value**	**Risk Parity**	**Halflife**	**Volume**
Annualized Mean Return in %	18.52	18.75	17.16	18.72	19.78
Annualized Standard Deviation in %	6.79	7.10	6.26	6.90	7.40
Sharpe Ratio	2.73	2.64	2.74	2.71	2.67
Win Rate in %	80.83	78.33	80.00	85.83	79.17
Maximum Drawdown in %	1.24	1.61	1.22	1.19	1.31

Table 10: Performance metrics comparison across different weighting methodologies for high-volume pairs trading portfolios during the testing period (2015-2024).

The strong performance of equal weighting across both training and testing periods confirms that when using high-volume stocks with optimized parameters, a straightforward allocation approach can deliver exceptional results. For simplicity and practical implementation, equal weighting is adopted for subsequent analysis.

5.5 Cost-Adjusted Return

The optimal parameter configuration from high-volume stocks with equal weighting demonstrated strong performance in both training and testing periods. However, real-world implementation requires consideration of transaction costs, which can significantly impact profitability. This analysis examines how different cost structures affect strategy performance during the training period (2005-2014) and later during testing period (2014-2024).

Transaction costs in pairs trading are notoriously difficult to capture as they vary significantly depending on trade size, investor type, the specific stocks being traded, and evolve considerably over time. In their comprehensive study covering the period 1963-2009, B. Do and Faff (2012) used 0.34% per trade as one of their transaction cost estimates. Building on this benchmark, this analysis examines a range of cost scenarios extending toward worst-case implementations to assess strategy robustness.

Two primary transaction costs affect pairs trading strategies: borrowing fees and trading commission fees. Borrowing fees represent the cost of shorting stocks, an essential component of pairs trading where one stock is bought while another is simultaneously borrowed and sold short. These fees typically range from 0.3% annually for highly liquid securities to 3% or more for less liquid stocks. Trading commission fees are charges incurred when executing trades, ranging from approximately 0.2% per trade for institutional investors to 2% for retail traders, reflecting different levels of market access and negotiating power.

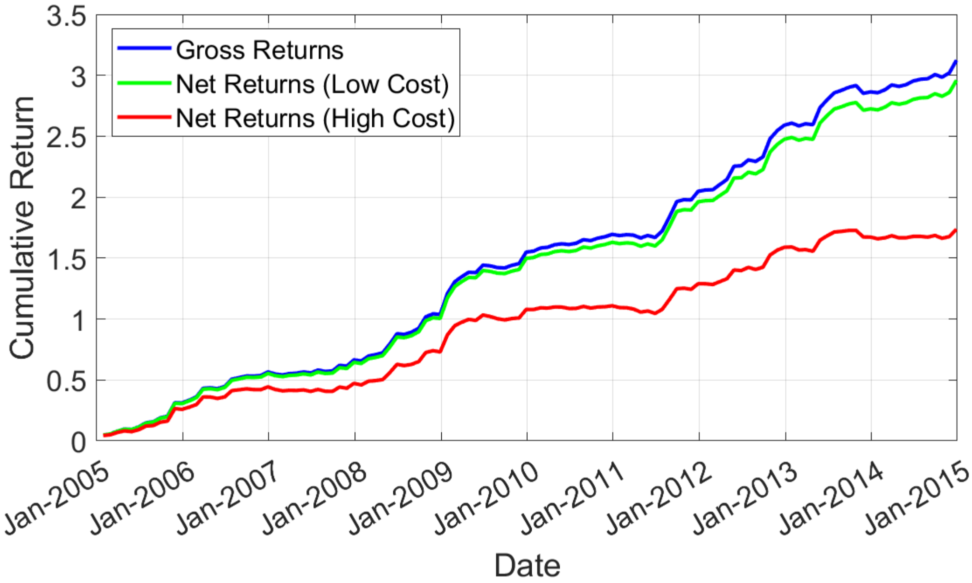


Figure 20: Cumulative returns comparison of gross returns versus cost-adjusted returns under low and high transaction cost scenarios for high-volume pairs trading portfolio during the training period (2005-2014).

Figure 20 illustrates the impact of these costs on cumulative returns, with the gross returns (blue line) representing the strategy without transaction costs. The low-cost scenario (green line) applies the lowest rates for both borrowing fees (0.3%) and trading commissions (0.2%), while the high-cost scenario (red line) incorporates the highest rates for both cost types (3% borrowing fees and 2% trading commissions). The most striking observation is the substantial divergence between the high-cost and low-cost scenarios, particularly following the 2008-2009 financial crisis, where the high-cost scenario captured significantly less of the strategy's potential.

For institutional investors with substantial capital and direct market access, the low-cost scenario represents a more realistic implementation environment, benefiting from lower borrowing fees on liquid securities and institutional-level commission rates. Conversely, the high-cost scenario better reflects the reality for retail investors or smaller institutional players who face higher borrowing costs and commission fees. Given that pairs trading strategies typically require significant capital and sophisticated risk management systems, the low-cost scenario likely represents the more realistic benchmark for evaluating the strategy's commercial viability.

The performance metrics in Table 11 reveal that under the low-cost scenario, the strategy remains highly attractive with only a slight reduction in annualized returns from 14.43% to 14.02% and a modest decrease in Sharpe ratio from 2.35 to 2.28. The win rate remains robust at 75% with minimal increase in maximum drawdown. However, the high-cost scenario demonstrates a decline in performance, with annualized returns dropping to 10.29% – still outperforming the market benchmark. The win rate falls to 64%, and maximum drawdown increases to nearly 3%, while cumulative returns plummet from 312% to 174%.

Metric	Gross Return			Net Return	
	High Vol. opt.	Market	SMB	Low Cost	High Cost
Annualized Return (%)	14.43	6.52	14.48	14.02	10.29
Annualized STD (%)	6.15	16.31	14.69	6.15	6.15
Sharpe Ratio	2.35	0.40	0.99	2.28	1.67
Win Rate (%)	75.00	59.17	62.50	75.00	64.17
Max Drawdown (%)	1.65	54.67	19.54	1.68	2.98
Cumulative Return (%)	312	67	283	296	174

Table 11: Performance metrics comparison of high-volume pairs trading portfolio under different transaction cost scenarios versus benchmark portfolios during the training period (2005-2014).

Despite this significant reduction, the high-cost scenario still delivers superior risk-adjusted performance compared to both benchmarks. The market achieved a lower annualized return (6.52%) with more than double the volatility (16.31%) and a devastating maximum drawdown of 54.67%, resulting in a much lower Sharpe ratio of 0.40. The SMB benchmark delivered comparable gross returns (14.48% annualized, 283% cumulative) but with substantial volatility (14.69%) and significant drawdown risk (19.54%), yielding a Sharpe ratio of 0.99. Importantly, these SMB figures represent gross returns and would likely deteriorate further after accounting for transaction costs and implementation frictions. Even under high transaction costs, the pairs trading strategy maintains superior risk control and more consistent performance than either benchmark, highlighting the strategy's inherent risk management advantages.

The testing period (2015-2024) shows similar cost impact patterns as the training period, with transaction costs reducing returns but maintaining superior risk-adjusted performance compared to the market. Figure 21 illustrates the cumulative returns under different cost scenarios, while Table 12 provides the corresponding performance metrics.

The gross returns (blue line) represent the idealized performance without transaction costs, achieving an impressive cumulative return of 515% over the decade. When implementing realistic low-cost assumptions (green line), the strategy maintains nearly all of its effectiveness, with cumulative returns of 490% and only a slight reduction in annualized returns from 18.52% to 18.1%. The Sharpe ratio remains exceptional at 2.67, with the win rate staying robust at 80%.

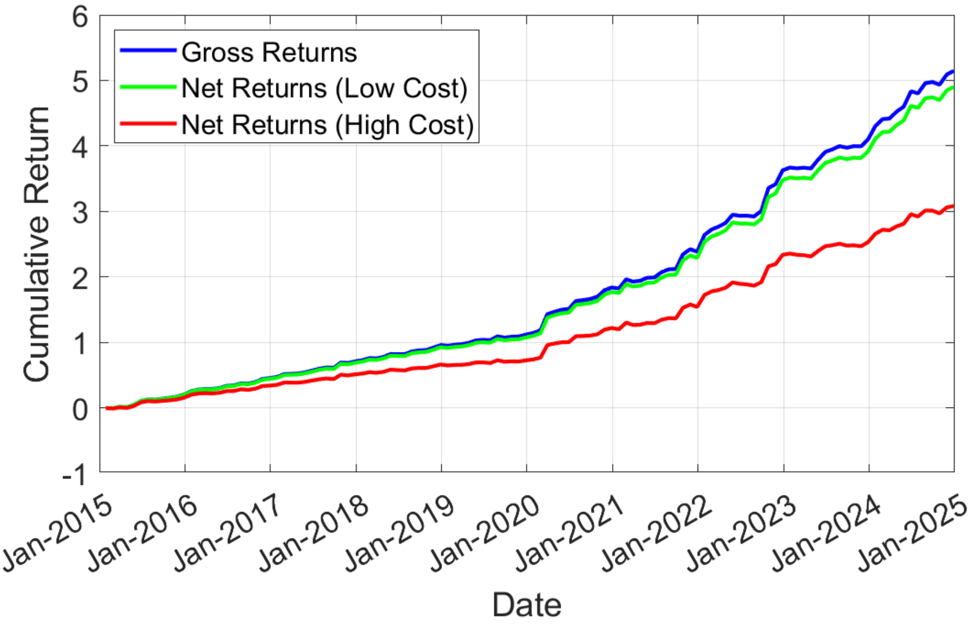


Figure 21: Cumulative returns comparison of gross returns versus cost-adjusted returns under low and high transaction cost scenarios for high-volume pairs trading portfolio during the testing period (2015-2024).

Even under the high-cost scenario (red line), the strategy maintains considerable profitability, with a cumulative return of 308% and an annualized return of 14.37%. While this represents a remarkable reduction compared to the gross return of 18.52%, it still outperforms both benchmarks on a risk-adjusted basis. The market benchmark delivered an annualized return of 8.78% but with more than double the volatility (15.08%) and a severe maximum drawdown of 25.89%, resulting in a much lower Sharpe ratio of 0.58.

Metric	Gross Return			Net Return	
	High Vol. opt.	Market	SMB	Low Cost	High Cost
Annualized Return (%)	18.52	8.78	9.96	18.10	14.37
Annualized STD (%)	6.79	15.08	14.61	6.79	6.79
Sharpe Ratio	2.73	0.58	0.68	2.67	2.12
Win Rate (%)	80.83	65.00	57.50	80.00	70.83
Max Drawdown (%)	1.24	25.89	17.93	1.27	1.69
Cumulative Return (%)	515	114	143	490	308

Table 12: Performance metrics comparison of high-volume pairs trading portfolio under different transaction cost scenarios versus benchmark portfolios during the testing period (2015-2024).

The SMB benchmark achieved similar returns (9.96% annualized, 143% cumulative) but also suffered from high volatility (14.61%) and substantial drawdown risk (17.93%), yielding a Sharpe ratio of only 0.68.

The high-cost scenario's Sharpe ratio of 2.12 still indicates strong risk-adjusted performance, significantly outpacing both benchmarks, with a solid win rate of 70.83% and well-controlled maximum drawdown of just 1.69%.

Comparing the cost impact across both periods reveals interesting insights. The strategy demonstrates greater resilience to transaction costs in the testing period than in the training period, with the high-cost scenario maintaining better relative performance. This suggests that the pairs identified during this period exhibited stronger mean-reversion tendencies, allowing the strategy to overcome cost hurdles more effectively.

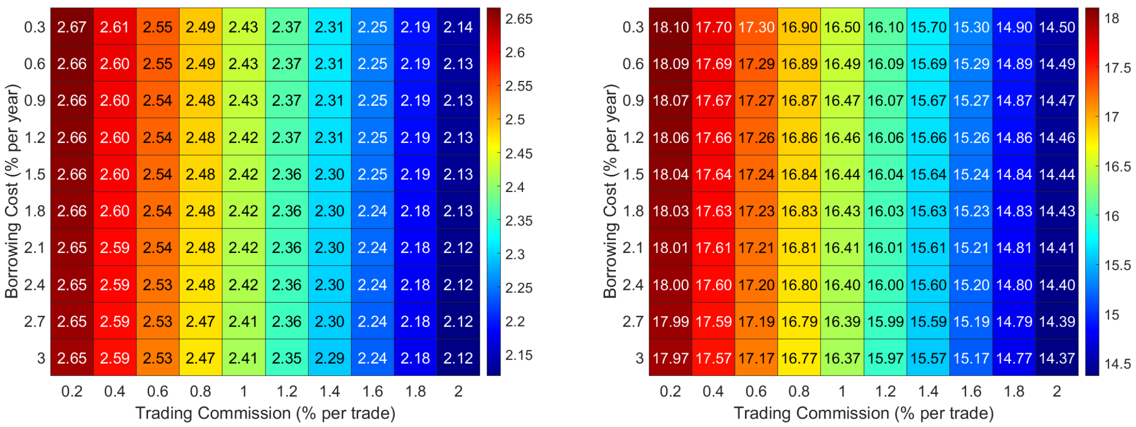
For the testing period, a cost sensitivity analysis was conducted to evaluate performance under different transaction cost scenarios. Figure 22 presents a detailed cost sensitivity analysis through two heatmaps showing cost-adjusted Sharpe ratio (left) and cost-adjusted annualized return (right) across different combinations of borrowing costs and trading commissions.

The heatmaps reveal several important insights about cost sensitivity. The Sharpe ratio (left panel) shows a gradient from red (highest values around 2.67) in the upper-left corner to blue (lowest values around 2.12) in the lower-right corner. This pattern indicates that both types of costs negatively impact performance, but with different degrees of influence. The horizontal bands display relatively consistent colors, while moving vertically produces more dramatic color changes. This visual pattern confirms that trading commission increases have a more substantial impact on performance than equivalent increases in borrowing costs.

The annualized return heatmap (right panel) exhibits a similar pattern but with more pronounced effects. Returns decline from approximately 18% in the most favorable cost scenario to 15% in the most expensive scenario. The rapid deterioration along the vertical axis (trading commission increases) compared to the more gradual changes along the horizontal axis (borrowing cost increases) further emphasizes the disproportionate impact of trading commissions on strategy profitability.

This asymmetric sensitivity to cost types origins from the strategy's fundamental mechanics. Borrowing costs are incurred continuously while holding short positions, but these positions typically remain open for relatively short periods (maximum one month under the optimal parameters). In contrast, trading commissions are incurred at both entry and exit for every trade, directly reducing profits regardless of how successful the trade might be. The frequency of trading (multiple positions opened and closed each month) amplifies the impact of commission costs.

These findings highlight two critical paths for improving pairs trading implementation. First, institutional investors with access to favorable commission structures maintain a substantial advantage in implementing such strategies. Second, strategy improvements that reduce turnover without sacrificing alpha generation could significantly improve the profitability. The cost sensitivity analysis ultimately demonstrates that pairs trading remains viable even under moderate cost scenarios, but implementation details become increasingly critical as costs rise. The strategy continues to deliver positive risk-adjusted returns across most examined cost structures, but the magnitude of outperformance depends heavily on an investor's ability to minimize transaction costs—particularly trading commissions.



(a) Sharpe ratio and varying cost parameters

(b) Annualized return and varying cost parameters

Figure 22: Sensitivity analysis of pairs trading performance to varying transaction cost parameters during the testing period (2015-2024)

5.6 Convergence within the Holding Period

An aspect of pairs trading strategy evaluation is understanding how frequently trades converge within the predetermined holding period. In the pairs trading framework, trades are closed under three conditions: when pairs converge by reaching their target exit thresholds, when stop-loss limits are triggered, or when the maximum holding period (one month) expires. This analysis examines the convergence rates across different volume configurations and time periods.

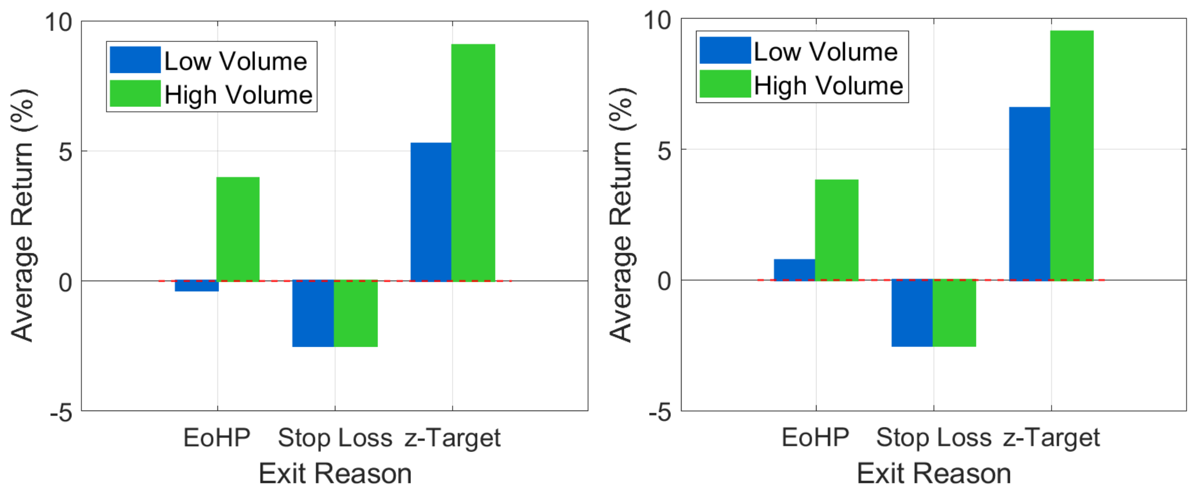
Metric	Period 1 (2005-2014)		Period 2 (2015-2024)	
	Low Volume	High Volume	Low Volume	High Volume
Average Target Exit Rate (%)	1.99	1.70	2.35	2.48
Std dev. Target Exit Rate (%)	1.77	1.94	3.04	2.76
Average Stop Loss Exit Rate (%)	59.75	66.55	61.78	64.04
Std dev. Stop Loss Exit Rate (%)	11.06	10.12	10.62	10.97
Average End of Holding Period Rate (%)	38.26	31.76	35.87	33.48
Std dev. End of Holding Period Rate (%)	11.13	10.06	11.04	10.97
Total Trades	32,220	32,387	28,697	31,214

Table 13: Distribution of exit reasons (target convergence, stop-loss trigger, and holding period expiration) for pairs trading strategies by volume.

The exit rate analysis in Table 13 reveals several important patterns in pairs trading behavior across different volume configurations and time periods. Target exit rates remain consistently low across all scenarios, ranging from 1.70% to 2.48%, indicating that pairs rarely converge to their target thresholds within the one-month holding period. However, this one-month holding period emerged from parameter optimization, suggesting that while longer holding periods might allow more pairs to converge, they would likely result in lower Sharpe ratios due to reduced risk-adjusted returns.

Stop-loss exits dominate across all configurations, accounting for approximately 60-67% of all trades. High-volume pairs show slightly higher stop-loss rates than low-volume pairs in both periods, suggesting that high-volume stocks may experience more frequent adverse price movements that trigger protective stops. The remaining 32-38% of trades reach the end of the holding period without triggering either exit condition, with low-volume pairs showing slightly higher rates of holding period expiration.

The accompanying Figure 23 provides crucial insights into the profitability of different exit reasons. High-volume pairs demonstrate superior performance across all exit reasons, with particularly strong returns when reaching target exits (approximately 9% average return) and positive returns even when stopped out at the end of holding period (EoHP). Low-volume pairs show more modest performance, with target exits generating around 5% returns and stop-loss exits producing small losses of -2.5%.



(a) Training period (2005-2014)

(b) Testing period (2015-2024)

Figure 23: Average returns of individual pairs by exit reasons across volume deciles and time periods, showing the profitability distribution of different trade termination conditions.

These findings have significant implications for pairs trading strategy design. Although stop-loss exits constitute the majority of trades, the strategy still performs well because the pairs that do converge (either through target exits or end-of-period exits) generate sufficiently high returns to compensate for the more frequent but smaller losses from stop-loss exits. This reinforces the importance of volume selection in pairs trading, as high-volume pairs not only provide better overall returns but also demonstrate more consistent profitability across all exit mechanisms, creating a more robust risk-return profile.

6 Discussion

This research systematically investigated interconnected questions about pairs trading effectiveness and optimal implementation in contemporary markets. The findings provide clear answers while revealing important insights about the strategy's reliability for modern investors and the mechanisms underlying its continued success.

The central research question examined whether pairs trading maintains its profitability in today's sophisticated market environment. The evidence overwhelmingly confirms that pairs trading remains highly effective, with the strategy delivering consistent positive returns across multiple evaluation metrics. Most remarkably, the strategy's performance actually improved in the testing period (2015-2024) compared to the training period (2005-2014), contradicting expectations that widely-known arbitrage strategies should experience diminishing returns over time as more participants exploit them.

The research revealed substantial differences in performance based on market microstructure characteristics, particularly trading volume. High-volume pairs consistently outperformed low-volume pairs across all evaluation metrics, achieving superior risk-adjusted returns and demonstrating higher return when converging towards the mean. This performance differential likely stems from enhanced liquidity facilitating more efficient price discovery and faster convergence to equilibrium relationships. However, an additional contributing factor may be the inherent similarity among stocks within the same volume decile. By constructing pairs from securities that share similar trading characteristics—whether heavily or lightly traded—the methodology naturally groups stocks with comparable investor behavior patterns. This similarity in trading dynamics could enhance the stability of price relationships and contribute to the superior performance observed in volume-segmented approaches. Furthermore, the low volatility observed in pairs trading portfolios likely results from the high similarity between paired stocks, which prevents them from diverging as dramatically as would occur in traditional long-short portfolios with securities not necessarily being co-integrated.

The systematic parameter optimization revealed that while pairs trading returns exhibit significant sensitivity to implementation choices, stable regions of superior performance exist. Notably, the optimized parameter configurations were remarkably similar between low and high volume deciles, suggesting consistent underlying market dynamics across different liquidity environments. The optimal configuration—2-year lookbook period in data, 1-month maximum holding period, entry z-score threshold of 1, exit threshold of 0.1, and stop-loss boundary of -2.5%—demonstrated robust performance across different market conditions. The superior performance of longer lookback periods paired with shorter holding periods reveals that because price relationships between stocks develop over recent timeframes, profitable convergence opportunities are relatively short-lived.

Contrary to expectations about temporal decay in arbitrage strategies, pairs trading effectiveness actually improved over time. The strategy demonstrated remarkable resilience during various market regimes, including the 2008 financial crisis and 2020 COVID-19 pandemic, consistently outperforming benchmark portfolios during periods when markets experienced significant volatility or stagnation. This sustained effectiveness suggests that either institutional constraints prevent full arbitrage of these opportunities or that market structural changes have enhanced rather than diminished the strategy's profitability.

The comparison of weighting methodologies revealed that while approaches do influence performance, the differences are modest compared to the impact of pair selection and parameter optimization. Equal weighting produced one of the highest risk-adjusted returns in most scenarios, with the relative similarity across schemes simplifying practical implementation and suggesting that sophisticated weighting methodologies may not justify their additional complexity.

Transaction cost analysis addressed whether realistic trading expenses eliminate theoretical profitability. The findings demonstrate that pairs trading remains profitable across most realistic cost scenarios, though with disproportionate sensitivity to trading commissions compared to borrowing costs. The substantial divergence between high-cost and low-cost scenarios highlights the importance of execution efficiency, with institutional investors benefiting significantly from favorable commission structures and access to liquid securities. Given that stop-loss exits constitute the majority of trades (approximately 60-67%), an interesting avenue for future research would be to investigate whether pairs destined for stop-loss exits could be predicted ex-ante, potentially allowing for selective trade filtering that could reduce turnover and associated trading costs.

The sustained effectiveness of pairs trading raises important questions about the generalizability of these results. While this study focused on S&P 500 stocks—representing highly liquid, well-established companies with extensive analyst coverage—the findings may not directly translate to other markets or asset classes. The volume-based segmentation approach that proved so successful here depends on having sufficient securities within each volume decile to form meaningful pairs, which may not be available in smaller or less liquid markets. Additionally, the optimal parameters identified may be specific to the institutional and technological environment of U.S. equity markets during the study period, suggesting that application to other markets might require recalibrating these parameters to match different market microstructures and trading dynamics.

The strategy's continued profitability despite widespread knowledge suggests that market frictions, implementation constraints, or structural factors prevent full convergence to theoretical efficiency. The low correlation with market returns and strong performance during stress periods highlight potential value as a diversification tool beyond standalone performance metrics. However, the cost sensitivity and requirement for sophisticated risk management systems may limit accessibility to institutional investors or individuals with substantial capital.

These findings collectively demonstrate that pairs trading remains viable within the S&P 500 universe, provided investors can access appropriate liquidity and cost structures. The research provides practical guidance on optimal implementation while highlighting continued market inefficiencies even within highly efficient markets. However, the extent to which these results generalize to other market segments, time periods, or geographical regions remains a question requiring further investigation.

7 Conclusion

This study investigated pairs trading strategies across market segments, parameters, weighting methods, and cost structures using U.S. equity data (2005-2024). Volume emerged as the dominant performance factor, with high-volume pairs consistently outperforming low-volume counterparts across multiple metrics, including superior returns when converging towards the historical relationship. The success of volume-based segmentation may stem partly from grouping stocks with similar trading characteristics, creating more stable price relationships. Notably, optimal parameter configurations were remarkably similar between volume deciles, demonstrating consistent underlying market dynamics across liquidity environments.

Contrary to efficient market predictions, performance improved in the testing period (2015-2024), suggesting structural market changes or institutional constraints have preserved the strategy's effectiveness despite widespread documentation. Different weighting methodologies produced similar results, with equal weighting combining simplicity and strong performance.

Transaction cost analysis revealed the strategy remains successful under institutional-level costs but deteriorates significantly with retail-level costs, with trading commissions having greater impact than borrowing fees. This study demonstrates that properly implemented pairs trading strategies continue to offer attractive risk-adjusted returns, particularly for institutional investors with favorable execution capabilities and cost structures.

8 Appendix

8.1 Procedure of Pairs Trading

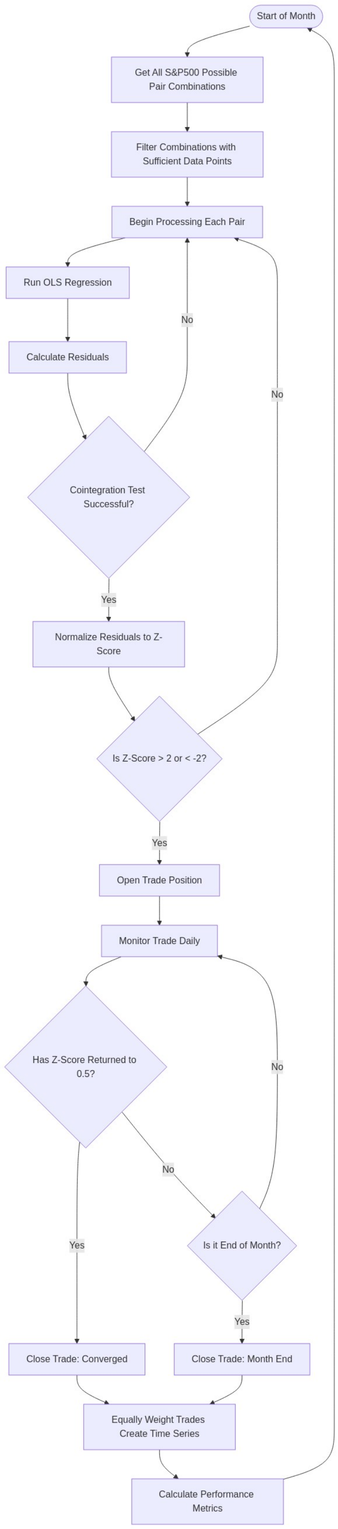


Figure 24: Flowchart for pairs trading

8.2 SMB portfolio based on S&P 500

Although SMB is well documented in the literature, it is presented here as a benchmark example for comparison purposes.

To demonstrate the effectiveness of the size effect, the analysis first examines S&P 500 companies from 2000 to 2004 as an illustrative example. Companies are sorted by market capitalization from smallest to largest, creating ten portfolios based on company size to track their performance. A clear pattern emerged across these portfolios: as company size increased, annualized returns consistently decreased, see Figure 25. This observed inverse relationship between company size and returns aligns with the well-documented size effect in financial literature and validates the rationale for constructing an SMB benchmark.

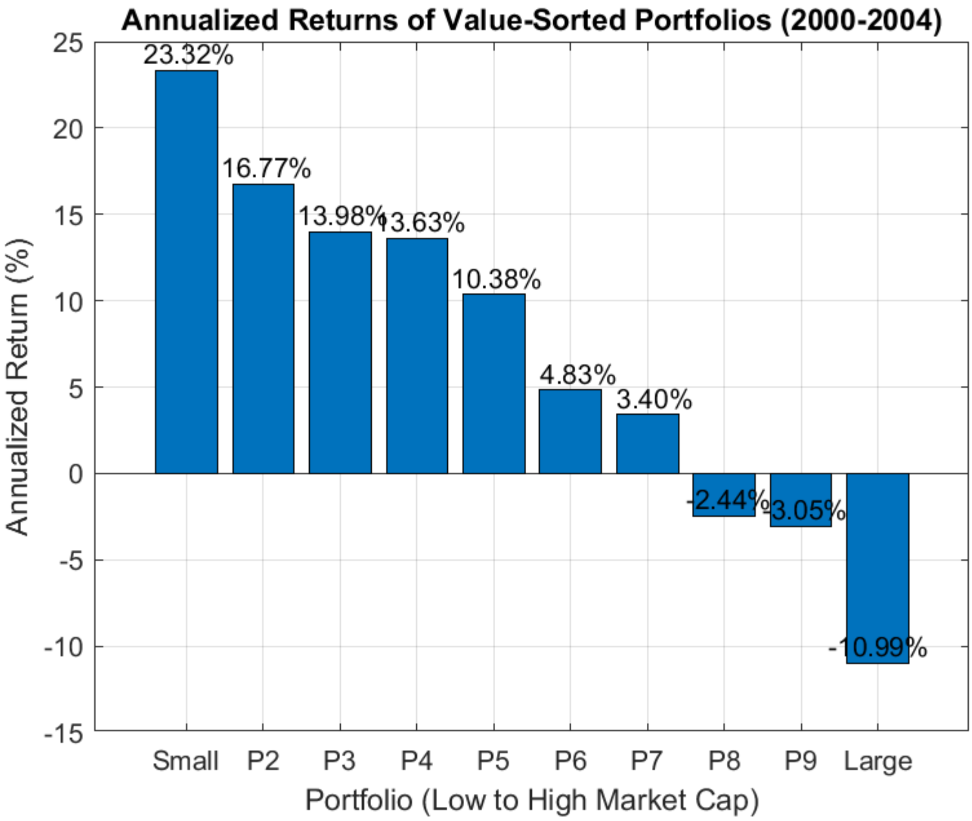


Figure 25: Annualized returns of value-sorted Portfolios (2000-2004)

8.3 Return Maximization

The parameter optimization process reveals striking differences between low-volume and high-volume stock pairs. Figure 26 illustrates the distribution of annualized mean returns across all 270 parameter combinations for both volume segments. The low-volume decile shows a distribution sharply concentrated around zero, indicating generally poor performance regardless of parameter settings. In contrast, the high-volume decile exhibits a wider distribution centered around 5.5% with an extended right tail reaching as high as 35% for optimal configurations. This disparity suggests that high-volume stocks perform better in pairs trading strategies.

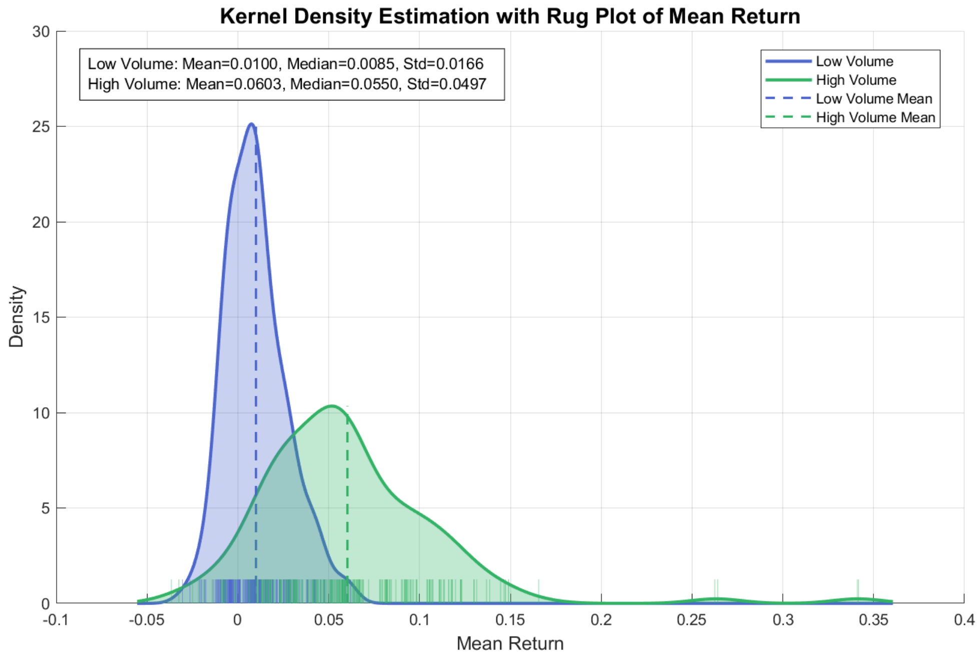


Figure 26: Distribution of annualized mean return in low and high volume deciles.

Examining specific parameters, Figure 27 plots exit z-score thresholds against annualized mean returns. Here again, the high-volume decile consistently achieves superior returns across different exit threshold values, while low-volume combinations cluster near zero regardless of the exit threshold selected. This pattern persists across other parameters, suggesting that stock volume characteristics dominate parameter selection effects for these segments.

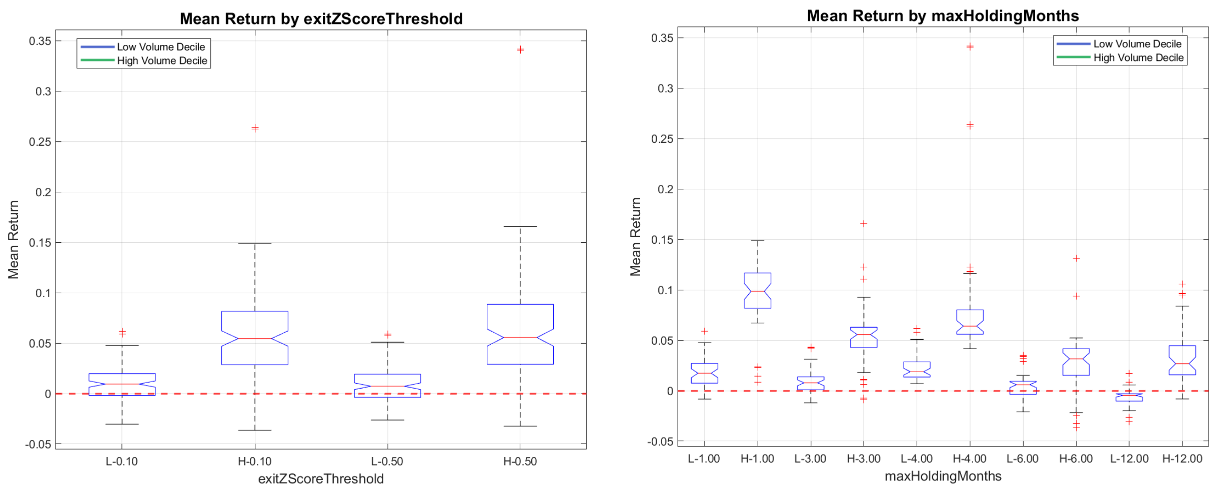


Figure 27: Annualized mean return in dependency of input parameters

Figure 27 further presents a particularly informative relationship between maximum holding periods and annualized returns. For high-volume pairs, a clear negative correlation emerges, as holding periods increase, returns systematically decrease. This pattern suggests that for high-volume stocks, price discrepancies correct relatively quickly, and extending holding periods merely dilutes returns by maintaining positions after convergence opportunities have dissipated. The optimal strategy for high-volume pairs appears to favor shorter holding periods that capitalize on rapid mean reversion before moving to new opportunities. Collectively, these findings challenge conventional wisdom that statistical arbitrage performs better in less liquid securities. Instead, the results demonstrate that high-volume stocks not only provide superior return potential but also respond more predictably to parameter adjustments. This may reflect more reliable mean-reverting behavior in liquid securities, potentially due to greater market efficiency in correcting temporary mispricing while avoiding the extended dislocations that can trap positions in less liquid pairs.

Lower Volume Decile

To ensure the robustness of the strategy and avoid overfitting, the dataset was divided into training (2005-2014) and testing (2015-2024) periods. Table 14 presents the optimized parameters and performance metrics for the lower volume decile when maximizing annualized returns. The optimal configuration features a short lookback period (2 years), medium holding period (4 months), high entry threshold (z-score of 3), tight exit threshold (0.1), and a moderate stop-loss (-5%). The performance shows some degradation in the testing period, with annualized return decreasing from 6.2% to 3.6%. However, risk characteristics remained stable across both periods, with nearly identical standard deviation (9% vs 9.1%) and maximum drawdown (12.8% vs 12.4%). The win rate showed only a modest decline from 55% to 52%. The cumulative return of 30.42% in testing is lower than the 68.16% during training.

Parameter/Metric	**Training (2005-2014)**	**Testing (2015-2024)**
Lookback Years	2	2
Max Holding Months	4	4
Entry Z-score	3	3
Exit Z-score	0.1	0.1
Stop-loss in %	- 5	- 5
Annualized Return in %	6.2	3.6
Annualized STD in %	9	9.1
Sharpe Ratio	0.69	0.40
Win Rate in %	55	52
Max Drawdown in %	12.8	12.4
Cumulative Return in %	68.16	30.42

Table 14: Input parameters when optimised towards highest annualized mean return in low volume decile

(a) Training period (2005-2014)

(b) Testing period (2015-2024)

Figure 28: Cumulative return over training and testing period

Upper Volume Decile

Table 15 presents the optimized parameters and performance metrics for the upper volume decile. The optimal configuration shares similarities with the lower volume decile in several parameters but employs a less stringent exit threshold (0.5) and tighter stop-loss (2.5%). The contrast between training and testing periods is stark. While the strategy generated exceptional returns during the training period (34.2% annualized return, 439% cumulative return), it failed to maintain this performance in the testing period, delivering negative returns (-1.12% annualized, -12.2% cumulative). The win rate declined from 45.8% to just 21%.

Parameter/Metric	**Training (2005-2014)**	**Testing (2015-2024)**
Lookback Years	2	2
Max Holding Months	4	4
Entry Z-score	3	3
Exit Z-score	0.5	0.5
Stop-loss in %	2.5	2.5
Annualized Return in %	34.2	-1.12
Annualized STD in %	71.7	6.8
Sharpe Ratio	0.48	-0.17
Win Rate in %	45.8	21
Max Drawdown in %	29.5	20
Cumulative Return in %	439	-12.2

Table 15: Input parameters when optimised towards highest annualized mean return in high volume decile

This dramatic performance reversal suggests that optimizing solely for maximum returns may lead to overfitting and poor out-of-sample performance. The strategy appears to have captured specific market anomalies in the training period that did not persist into the testing period. These results indicate that a more balanced optimization approach focusing on risk-adjusted returns rather than raw returns might produce more robust and sustainable performance across different market regimes.

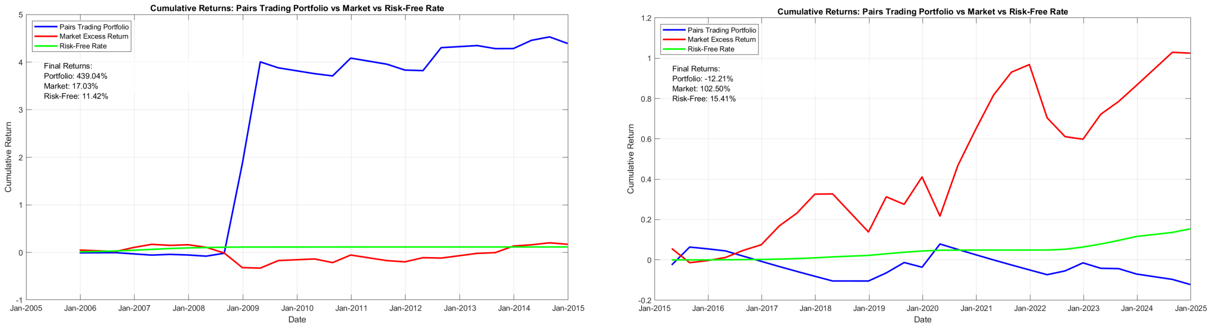


Figure 29: Cumulative return during training and testing period

9 Bibliography

References

Avellaneda, M., & Lee, J.-H. (2010). Statistical arbitrage in the us equities market. *Quantitative Finance*, *10*(7), 761–782.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18.

Bowen, D. A., & Hutchinson, M. C. (2016). Pairs trading in the uk equity market: Risk and return. *The European Journal of Finance*, *22*(14), 1363–1387.

Chan, E. P. (2013). *Algorithmic trading: Winning strategies and their rationale*. John Wiley & Sons.

Chen, H., Chen, S., Chen, Z., & Li, F. (2019). Empirical investigation of an equity pairs trading strategy. *Management Science*, *65*(1), 370–389.

CRSP. (2025). CRSP US Stock Database [Accessed: February 15, 2025]. *Booth School of Business, The University of Chicago*. Retrieved February 15, 2025, from https://www.crsp.org/products/research-products/crsp-us-stock-databases

Do, B., & Faff, R. (2012). Are pairs trading profits robust to trading costs? *Journal of Financial Research*, *35*(2), 261–287.

Do, H. X., & Brooks, R. D. (2015). Cointegration and relative market efficiency: An analysis of asian emerging markets. *The Quarterly Review of Economics and Finance*, *58*, 45–53.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Figuerola-Ferretti, I., Paraskevopoulos, I., & Tang, T. (2018). Pairs-trading and spread persistence in the european stock market. *Journal of Futures Markets*, *38*(9), 998–1023.

Flori, A., & Regoli, D. (2021). Revealing pairs-trading opportunities with long short-term memory networks. *European Journal of Operational Research*, *295*, 772–791.

French, K. R. (2025). Data Library [Accessed: April 15, 2025]. *Tuck School of Business at Dartmouth*. Retrieved April 15, 2025, from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

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.

Han, C., He, Z., & Toh, A. J. (2023). Pairs trading via unsupervised learning. *European Journal of Operational Research*, *307*, 929–947.

Hong, S., & Hwang, S. (2023). In search of pairs using firm fundamentals: Is pairs trading profitable? *The European Journal of Finance*, *29*(5), 508–526.

Jacobs, H., & Weber, M. (2015). On the determinants of pairs trading profitability. *Journal of Financial Markets*, *23*, 75–97.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2-3), 231–254.

Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys*, *31*(2), 513–545.

Leung, T., & Li, X. (2015). *Optimal mean reversion trading: Mathematical analysis and practical applications.* World Scientific.

Reinganum, M. R. (1981). Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. *Journal of Financial Economics*, *9*(1), 19–46.

Sarmento, S. M., & Horta, N. (2020). Enhancing a pairs trading strategy with the application of machine learning. *Expert Systems with Applications*, *158*, 113490.

Tadi, M., & Davallou, M. (2021). Evaluation of dynamic cointegration-based pairs trading strategy in the cryptocurrency market. *arXiv preprint arXiv:2109.10662*.