New York University Stern School of Business | Robo Advisors and Systematic Trading

Designing Alpha: Machine Learning & Stochastic Models for Systematic Trading

Komal Niraula, Student ID: kn2505

New York, 15th May 2025

https://github.com/komalniraula/ml-quant-eval

Abstract

This paper evaluates a mean-reversion pairs trading strategy using the Ornstein-Uhlenbeck (OU) process to model relationships between stock prices. The methodology incorporates K-means clustering for fundamental peer grouping and GARCH(1,1) for volatility forecasting, using data from CRSP and Compustat. The capital allocation approach invests a fresh $1 billion quarterly, distributing among the pairs using inverse volatility. Results show that an OU-based z-score method with a threshold of 1.5 and correlation filter of 0.9 achieves the highest in-sample Sharpe ratio (0.53) and Sortino ratio (0.94), despite a slightly negative CAPM alpha (-0.04).

The strategy maintains robustness in out-of-sample testing (2022-2024) with a Sharpe ratio of 0.36 and improved CAPM alpha of 0.002. The findings highlight an important trade-off between risk-adjusted returns and market-neutrality in pairs trading strategies. While higher alpha values were achievable using classical z-score methods, the OU-based approach delivered superior risk-adjusted performance with significantly lower return volatility. Notably, the OU-based method generated far fewer trades (739) compared to classical z-score approaches (3,880), suggesting greater signal precision.

Contents

1 Introduction . . . 2
2 Challenges in Optimization for Systematic Trading . . . 3
3 Conceptual Framework . . . 3
3.1 Statistical Arbitrage and Mean-Reversion Strategies . . . 3
3.2 Growth and Value Portfolio Focus . . . 3
3.3 Integrating Machine Learning and Quantitative Finance . . . 4
3.4 Portfolio Rebalancing Frequency . . . 7
3.5 Capital Allocation and Risk Management Framework . . . 7
4 Methodology . . . 7
4.1 Data Sources and Collection . . . 7
4.2 Data Cleaning and Preprocessing . . . 12
4.3 Sample Selection and Portfolio Formation . . . 12
4.4 Volatility Estimation Using GARCH Models . . . 14

4.5 Performance Evaluation Metrics . . . 17
4.6 Software and Tools . . . 17
5 Results and Analysis . . . 17
5.1 In-Sample Performance (2015–2021) . . . 18
5.2 Out-of-Sample Performance (2022–2024) . . . 18
5.3 Final Strategy Performance Comparison . . . 19
6 Conclusion . . . 20
7 Future Work . . . 21
References . . . 22

1 Introduction

In modern financial markets, the pursuit of alpha generation through market-neutral strategies has gained significant traction. Among these, pair trading, a form of statistical arbitrage, remains one of the most widely adopted strategies due to its intuitive reliance on mean reversion and relative pricing inefficiencies. Traditional approaches to pairs trading often rely on historical price correlations; however, these methods may overlook the underlying fundamental relationships that drive long-term asset value and fail to account for changing market volatility.

This study examines the potential for enhancing the effectiveness of pairs trading strategies by integrating machine learning techniques with traditional quantitative models. To implement this integration, peer portfolios are first constructed using the K-Means clustering algorithm applied to company's financial reports. Within these portfolios, pairs are selected using statistical tests such as correlation analysis, and cointegration testing with the Augmented Dickey-Fuller (ADF) stationarity test.

To improve signal reliability and account for time-varying market risk, this study incorporates the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model to estimate dynamic volatility. These volatility estimates are used to control market exposure and minimize risk. Trading signals are refined using the Ornstein-Uhlenbeck (OU) process, which models mean-reversion behavior and generates trading signals through OU-based z-scores. This approach provides a theoretically sound and statistically robust measure of how far asset prices have deviated from their long-term equilibrium.

The analysis focuses exclusively on growth and value portfolios, categorized using the Fama-French Book-to-Market (B/M) ratio framework. Additionally, the strategy design incorporates practical budget and liquidity constraints, ensuring that the backtesting environment realistically reflects capital deployment limitations and market frictions.

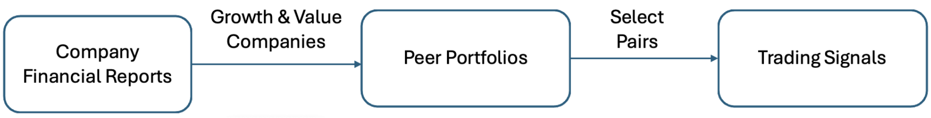


Figure 1: Process Flow Diagram for Pairs Trading Strategy

The primary objective of this research is to evaluate whether portfolios formed based on fundamental similarity and filtered through rigorous statistical techniques can consistently deliver statistically significant alpha. Beyond assessing the profitability of the proposed machine learning and stochastic modeling framework, this study specifically focuses on identifying the optimal threshold parameters, such as correlation filters and z-score levels, that govern trade execution decisions. The performance of the proposed strategy is evaluated through both in-sample (2015–2021) and out-of-sample (2022–2024) testing, providing a comprehensive assessment of its robustness and profitability under real-world constraints.

2 Challenges in Optimization for Systematic Trading

A critical challenge in implementing mean-reversion trading strategies lies in determining the optimal set of thresholds that effectively balance trade frequency and profitability. Poor selection of z-score thresholds or correlation filters can result in overtrading, weak signal quality, and ultimately, reduced returns. Traditional strategies often rely on static thresholds derived from historical averages; however, such approaches fail to account for dynamic market conditions and evolving volatility regimes.

This study addresses these challenges by systematically evaluating the impact of different threshold parameters within a robust framework that integrates machine learning and stochastic modeling techniques. Specifically, K-Means clustering) is employed to form peer groups based on companies' financial reports, ensuring relevant and dynamic portfolio construction. Additionally, advanced stochastic models such as GARCH and the Ornstein-Uhlenbeck process are incorporated to improve volatility estimation and return forecasting. By combining these approaches, this research aims to enhance the reliability of trading signals and optimize the overall performance of systematic trading strategies.

3 Conceptual Framework

3.1 Statistical Arbitrage and Mean-Reversion Strategies

Statistical arbitrage strategies aim to exploit temporary mispricings in financial markets while maintaining a market-neutral stance. A key foundation of such strategies is the concept of mean-reversion, which suggests that asset prices deviate from their historical averages but eventually revert back over time. This behavior is often attributed to overreaction and correction patterns in investor behavior. Identifying and exploiting these price deviations enables traders to generate alpha without assuming broad market risk.

3.2 Growth and Value Portfolio Focus

This study focuses exclusively on companies classified into **Growth and Value** portfolios, following the methodology introduced by **Fama and French** (Fama and French 1992) in their Three-Factor Model. The classification is based on the **Book-to-Market (B/M)** ratios. Companies falling in the **bottom 30%** of the B/M distribution within each quarter are classified as Growth firms, while those in the **top 30%** of the B/M distribution are classified as Value firms. This dynamic quarterly classification ensures that the portfolios remain responsive to the latest market conditions and financial performance of the companies under consideration.

3.3 Integrating Machine Learning and Quantitative Finance

3.3.1 Machine Learning Model for Peer Portfolio

In this research, the **K-Means Clustering** algorithm is employed to form localized peer groups based on companies' financial reports. Unlike traditional fixed-sector classifications, K-Means enables the creation of more relevant comparison groups that reflect the latest firm performance.

Given a stock with feature vector $\mathbf{x}$, the K-Means algorithm partitions the universe of stocks into k distinct clusters by minimizing the within-cluster sum of squared distances:

$$\min_{C} \sum_{i=1}^{k} \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

where:

- C_i is the set of stocks assigned to cluster i.
- $\boldsymbol{\mu}_i$ is the centroid (mean vector) of cluster i.
- $\mathbf{x}$ is the feature vector of a stock (including *Book-to-Market Ratio (B/M)*, *Return on Assets (ROA)*, and stability measures of profitability and revenue growth (Table 4)).

Clusters are formed on a quarterly basis using standardized fundamental features, ensuring that peer groups reflect the most recent financial performance of firms.

3.3.2 Quantitative Models for Signal Generation

To improve the robustness of trading signals, the following quantitative models are integrated:

- **GARCH (Generalized Autoregressive Conditional Heteroskedasticity)**: To overcome the limitations of rolling volatility, GARCH(1,1) is used to model time-varying volatility (Bollerslev 1986), allowing dynamic risk assessment. The conditional variance σ_t^2 is calculated as:

  $$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

  where:

  - The parameters α and β are estimated using Maximum Likelihood Estimation (MLE). This method identifies the values of α and β that maximize the probability of observing the historical return series. Once estimated, these parameters are used to forecast future volatility.
  - ω is the long-term average variance component, representing the baseline volatility level toward which the process reverts when no recent shocks are present.
    * It is estimated using past return data (this report uses the past 50 days of returns) through Maximum Likelihood Estimation (MLE), which identifies the parameter values that maximize the likelihood of observing the historical return series.
  - α measures the impact of recent shocks (short-term volatility reaction).

    * The term ϵ_{t-1}^2 represents the squared return shock from the previous day and is calculated as:

$$\epsilon_{t-1} = r_{t-1} - \mu \quad \text{and} \quad \epsilon_{t-1}^2 = (r_{t-1} - \mu)^2$$

      where r_{t-1} is the actual return of the previous day and μ is the expected mean return.
    * A higher value of α increases the weight assigned to recent large return shocks, causing today's forecasted variance to rise more sharply.
  - β captures the persistence of volatility (how long elevated volatility remains after a shock).
    * The term σ_{t-1}^2 is the forecasted variance from the previous day.
    * A higher β value means that past volatility has a stronger influence, and volatility remains elevated or suppressed for longer periods after a shock.
    * This persistence is estimated based on the variance patterns observed in the recent 30-day return window.
  - The sum $\alpha + \beta$ reflects the overall persistence of volatility in the time series.
    * If $\alpha + \beta \approx 1$, volatility is highly persistent, and shocks take longer to dissipate.
    * If $\alpha + \beta < 1$, volatility mean-reverts more quickly to its long-term average.

- **Ornstein-Uhlenbeck (OU) Process**: The OU process (Uhlenbeck and Ornstein 1930) is employed to model mean-reverting behavior of stock returns. Its discrete-time forecast formula is:

$$\mathbb{E}[X_{t+1}] = X_t e^{-\theta} + \mu(1 - e^{-\theta})$$

  where:
  - X_t is the current return,
  - θ is the speed of mean reversion (estimated using OLS regression: $\theta = -\ln(\beta)$),
  - μ is the long-term mean level of returns, calculated as $\mu = \frac{c}{1-\beta}$,
  - β and c are obtained from the regression equation $X_t = \beta X_{t-1} + c$.

- **Augmented Dickey-Fuller (ADF) Test**: The ADF test is applied to verify the stationarity of price spreads, ensuring that the selected groups exhibit mean-reverting characteristics. The test is based on the regression:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p} + \epsilon_t$$

  where the null hypothesis $H_0 : \gamma = 0$ indicates the presence of a unit root (non-stationary). Rejecting H_0 confirms stationarity and potential mean-reversion.

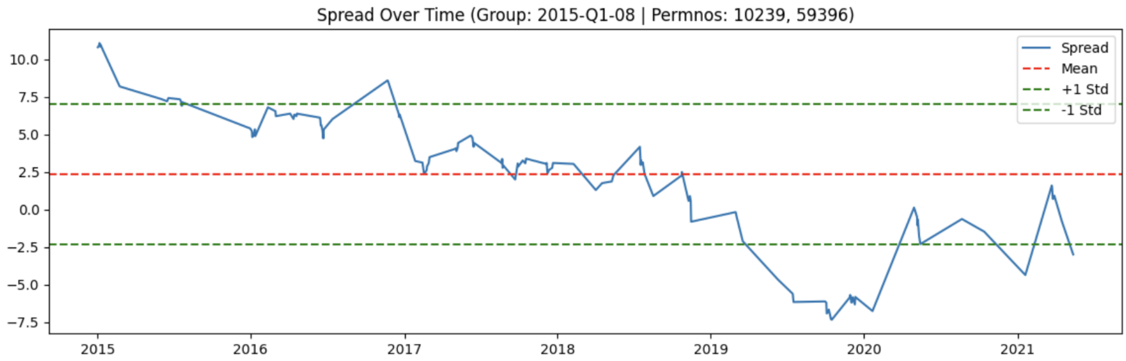


Figure 2: Spread Between Paired Stocks Showing Mean-Reversion Behavior with Mean (Red) and Standard Deviation Bands (Green)

3.3.3 Z-Score for Trading Signals

Trading signals in this study are generated using the standardized Z-score, which measures how far the return of one stock deviates from another. Under the assumption of mean reversion, a large deviation is expected to correct over time, creating profitable trading opportunities. Two approaches are implemented to calculate the Z-difference.

BLACK and WHITE are used as illustrative stock examples to explain the methodology.

1. **Classical Standard Deviation-Based Z-Diff:**
   This approach normalizes the current return by historical volatility without adjusting for the historical mean. It captures how large the return is relative to its typical variation.

$$Z_{\text{BLACK}} = \frac{R_{\text{BLACK}}}{\sigma_{\text{BLACK}}}, \quad Z_{\text{WHITE}} = \frac{R_{\text{WHITE}}}{\sigma_{\text{WHITE}}}$$
$$Z_{\text{DIFF}} = Z_{\text{BLACK}} - Z_{\text{WHITE}}$$

   Where:

   - R_{BLACK}, R_{WHITE}: Returns over a specific period that could be 5/10/15 day horizon.
   - σ_{BLACK}, σ_{WHITE}: Historical standard deviations over the lookback window.

   A positive Z_{DIFF} suggests that BLACK is overperforming relative to WHITE, indicating a potential short position on BLACK and a long position on WHITE. Conversely, a negative Z_{DIFF} signals the opposite.

2. **Ornstein-Uhlenbeck (OU)-Based Z-Diff:**
   This approach explicitly incorporates mean reversion by considering the deviation from the OU-predicted long-term equilibrium level.

$$Z_{\text{BLACK, OU}} = \frac{R_{\text{BLACK}} - \hat{\mu}_{\text{BLACK}}}{\sigma_{\text{BLACK}}}, \quad Z_{\text{WHITE, OU}} = \frac{R_{\text{WHITE}} - \hat{\mu}_{\text{WHITE}}}{\sigma_{\text{WHITE}}}$$
$$Z_{\text{DIFF, OU}} = Z_{\text{BLACK, OU}} - Z_{\text{WHITE, OU}}$$

   Where:

   - $\hat{\mu}_{\text{BLACK}}$, $\hat{\mu}_{\text{WHITE}}$: OU-predicted long-term mean returns for BLACK and WHITE.

- σ: Volatility estimate based on the volatility predicted by GARCH.

This method better captures the expected reversion behavior, as it directly measures how far each stock's return deviates from its equilibrium value predicted by the OU process.

In the in-sample period, various Z-score thresholds is tested for both the Classical and OU-Based methods to identify optimal entry and exit signals. The performance of each method under different thresholds is compared, and the best-performing method and thresholds are validated on out-of-sample data.

3.4 Portfolio Rebalancing Frequency

Given the dynamic nature of financial markets, it is crucial to periodically reassess and rebalance portfolios. This study adopts a quarterly rebalancing approach to incorporate updated financial data and ensure that the strategy remains responsive to changing market conditions.

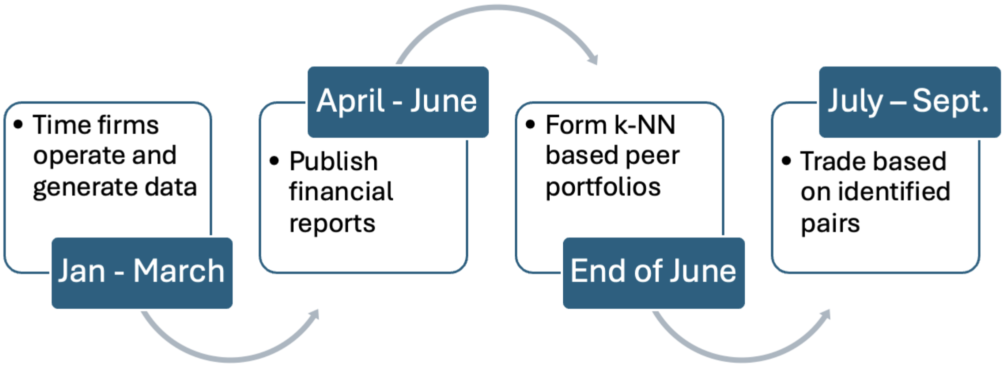


Figure 3: Quarterly Process Timeline (Example dates are shown to illustrate sequence)

As it typically takes time for companies to prepare and publish their financial reports, the financial data from one quarter is only available in the subsequent months.

3.5 Capital Allocation and Risk Management Framework

Effective capital allocation is essential for optimizing portfolio returns and managing risk. The study employs inverse volatility weighting to allocate more capital to less volatile pair, balancing expected returns against potential risks.

Transaction and financing costs are integrated into the allocation process to ensure realistic performance evaluation. The strategy also maintains strict budget and liquidity constraints to control overall exposure and minimize market impact.

4 Methodology

4.1 Data Sources and Collection

This study utilizes multiple financial data sources, all collected from Wharton Research Data Services (WRDS):

- **CRSP (Center for Research in Security Prices):** Provides daily stock prices and returns data.
- **Compustat:** Supplies quarterly firm-level fundamental data, including financial ratios and accounting variables.
- **Federal Reserve Economic Data (FRED):** Provides the daily Federal Funds Rate used for financing cost calculations.

The datasets are merged using unique firm identifiers (`permno`) and CRSP-Compustat Link Table.
Lookahead bias is carefully omitted in every step of the data preparation process, ensuring that information available only up to **t-1** is used to generate signals for **t (today)**.

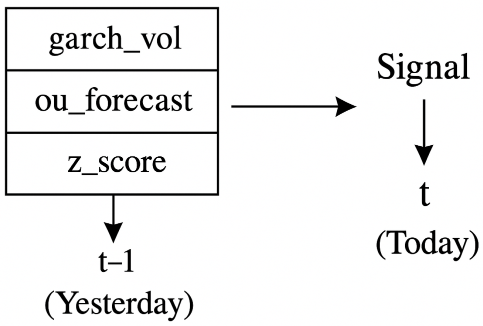


Figure 4: Illustration of avoiding Lookahead Bias

For GARCH volatility estimation and OU parameters, only returns up to **t-1** are used to forecast volatility and expected mean reversion behavior for the day t.

4.1.1 Variables for Identifying Growth and Value Portfolios

Table 1: Variables Used for Growth and Value Classification

Variable	Source Table	Description / Usage	Transformations
prc	CRSP (dsf)	Daily stock price	Absolute value taken
shrout	CRSP (dsf)	Shares outstanding	Converted to thousands
pstkrq	Compustat	Preferred Stock (Primary)	Fallback if missing
pstknq	Compustat	Preferred Stock (Secondary)	Used if pstkrq missing
seqq	Compustat	Stockholders' Equity (Primary)	Fallback applied if missing
atq	Compustat	Total Assets	Used for fallback in SEQ calculation
ltq	Compustat	Total Liabilities	Used for fallback in SEQ calculation
ceqq	Compustat	Common Equity	Used for fallback in SEQ calculation
pstkq	Compustat	Preferred Stock (Alternative)	Used for fallback in SEQ calculation
txditcq	Compustat	Deferred Taxes	Used for fallback in SEQ calculation

Table 2: Calculated Metrics for Growth and Value Classification

Metric	Formula	Description	Reference
Preferred Stock (PS)	PSTKRQ $\rightarrow$ PSTKNQ $\rightarrow$ $\ldots$	Fallback logic: Use primary preferred stock, else secondary, else 0.	Daniel et al. (1997)
SEQ (Stockholders' Equity)	SEQQ $\rightarrow$ (ATQ $-$ LTQ) $\rightarrow$ (CEQQ + PSTKQ + TXDITCQ)	Fallback logic for Stockholders' Equity if SEQQ is missing.	Cohen & Frazzini (2003)
Book Equity (BE)	SEQ $-$ PS	Book Value of Equity.	Fama & French (1992)
Market Equity (ME)	prc $\times$ shrout	Market Capitalization.	Banz (1981)
B/M Ratio	BE $\times$ 1000/ME	Book-to-Market Ratio for classification.	Rosenberg et al. (1985)

4.1.2 Variables Used for K-Means Clustering

Table 3: Variables Used for Fundamental Feature Calculation

Variable	Source Table	Description / Usage	Transformations
niq	Compustat	Net Income (Quarterly)	-
atq	Compustat	Total Assets (Quarterly)	-
oiadpq	Compustat	Operating Income After Depreciation	-
saleq	Compustat	Total Sales / Revenue	-
cogsq	Compustat	Cost of Goods Sold	-
capxq	Compustat	Capital Expenditures	-

Table 4: Fundamental Metrics Calculated for Clustering

Metric	Formula	Description	Reference
ROA	$\frac{\text{niq}}{\text{atq}}$	Return on Assets. Profitability measure relative to total assets.	Piotroski (2000)
Operating Margin	$\frac{\text{oiadpq}}{\text{saleq}}$	Efficiency of core business operations.	Novy-Marx (2013); Wahab et al. (2024)
Gross Margin	$\frac{\text{saleq}-\text{cogsq}}{\text{saleq}}$	Profit after production costs.	Piotroski (2000); Wahab et al. (2024)
Revenue Growth	% Change in saleq	Quarterly revenue growth percentage.	Li & Froot et al. (2016
Capex Intensity	$\frac{\text{capxq}}{\text{atq}}$	Capital investment relative to total assets.	Cooper et al. (2008)
ROA Stability	Std Dev of ROA over 8 Quarters	Measures consistency of profitability.	Li & Mohanram (2018)
Revenue Growth Stability	Std Dev of Revenue Growth over 8 Quarters	Measures volatility of revenue growth.	Li & Mohanram (2018)

Table 5: CRSP Variables for Trading Strategies

Variable	Source (CRSP)	Description	Transformations
prc	dsf	Adjusted Closing Price	Absolute Value
shrout	dsf	Shares Outstanding	In Thousands
retx	dsf	Return (Excluding Dividends)	-
vol	dsf	Trading Volume	Divided by 100
cfacpr	dsf	Price Adjustment Factor	Used for price adjustment
adj_prc	Calculated	Price adjusted for stock splits and dividends	prc/cfacpr

4.1.3 Federal Funds Rate Data for Financing Costs

Table 6: Federal Reserve Variables Used for Financing Cost Calculations

Variable	Source (FRED)	Description	Transformations
dff	frb.rates_daily	Daily Effective Federal Funds Rate (Discounted)	Used as fallback for missing effr
effr	frb.rates_daily	Effective Federal Funds Rate	Primary rate used for financing calculations
fed_funds_rate	Calculated	Final Federal Funds Rate for Strategy Calculations	fed_funds_rate = effr $\rightarrow$ dff (Fallback logic)

4.2 Data Cleaning and Preprocessing

Before any modeling and analysis, the merged datasets undergo rigorous preprocessing to ensure data quality and consistency:

- **Missing Values:** Missing financial variables are handled using fallback logic as shown in Tables 2 and 6. For key variables used in K-Means clustering, a two-stage hybrid imputation method is applied:
  1. **Time-Series Interpolation:** Missing values within each firm (`permno`) are filled using linear interpolation based on the firm's historical data.
  2. **Cross-Sectional Extrapolation:** For remaining missing values, the method estimates average market-wide percentage changes between periods and applies them to firms with incomplete data, ensuring values reflect broader market movements rather than static averages.

  Any remaining missing data after these steps are excluded from the analysis before portfolio formation and backtesting.

- **Outlier Treatment:** Extreme values in financial ratios and returns are winsorized at the 1st and 99th percentiles to reduce the influence of outliers on portfolio formation and trading strategies.

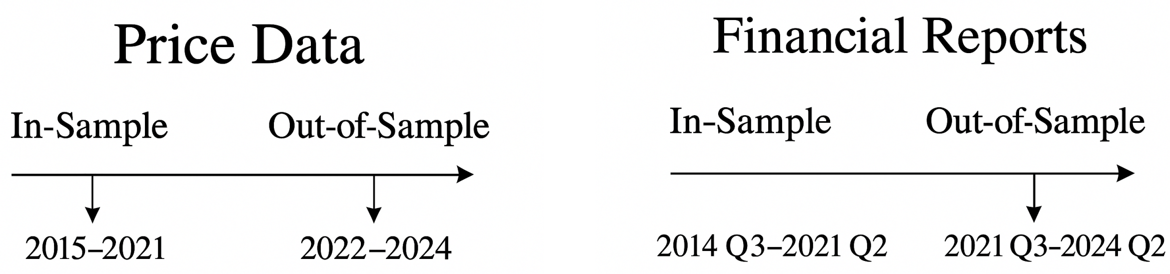
- **Feature Scaling:** For clustering, all fundamental features are standardized using Z-score normalization to ensure comparability across variables with different scales.

4.3 Sample Selection and Portfolio Formation

This study evaluates the effectiveness of mean-reversion strategies within portfolios classified as Growth and Value, based on the Fama-French Book-to-Market (B/M) ratio framework. While this classification provides a broad cross-section of the market, covering diverse firm characteristics and market behaviors, the investment universe is further narrowed through a systematic, data-driven peer selection process.

Peer portfolios are constructed quarterly using the K-Means clustering algorithm applied to selected fundamental variables. This approach forms niche peer groups based on the latest financial performance, ensuring that paired companies are fundamentally similar. The number of clusters is set to **11**, corresponding to the sectors defined by the **Global Industry Classification Standard (GICS)**, thereby maintaining sectoral consistency in peer grouping.

The analysis is restricted to companies listed on the three major U.S. stock exchanges—**NYSE, NASDAQ, and AMEX**. To maintain statistical reliability and avoid unstable pair selections, only clusters containing at least 30 unique securities are considered. Pair selection is then performed within each qualified cluster, preserving the fundamental similarity between the paired stocks. The selection is further evaluated through a dedicated in-sample (2015–2021) model development phase and an out-of-sample (2022–2024) validation period to ensure robustness and real-world applicability.



(a) Price Data Sample Timeline

(b) Reports Data Sample Timeline

Figure 5: Timeline Representation for In-Sample and Out-of-Sample Data Splits

Pair Selection Process:
Pair selection follows a two-stage filtering approach:

- **Stage 1: Correlation Analysis** Historical price correlations are computed to identify stock pairs exhibiting strong short-term co-movement. Only pairs with a correlation coefficient above **0.5** are retained for further analysis, which is further tested in in-sample method.
- **Stage 2: Cointegration Test (Engle-Granger with ADF)** For the filtered pairs, the cointegration is applied with the ADF test on the residuals. Only pairs with a **cointegration p-value below 0.05** are selected for trading. This threshold is kept fixed throughout the in-sample and out-of-sample evaluations due to time constraints and also as it represents a widely accepted standard for statistical significance.

Half-Life Filter: To ensure practical tradability and effective mean-reversion behavior, the half-life of mean reversion for each spread is estimated using the Ornstein-Uhlenbeck (OU) process. Pairs with a half-life exceeding **20 days** are excluded, as they exhibit weak or slow mean-reversion dynamics, making them less attractive for short- to medium-term trading.

Liquidity Constraints: The number of shares traded is limited to a maximum of **10% of a stock's 20-day average trading volume**, ensuring sufficient liquidity for efficient trade execution.

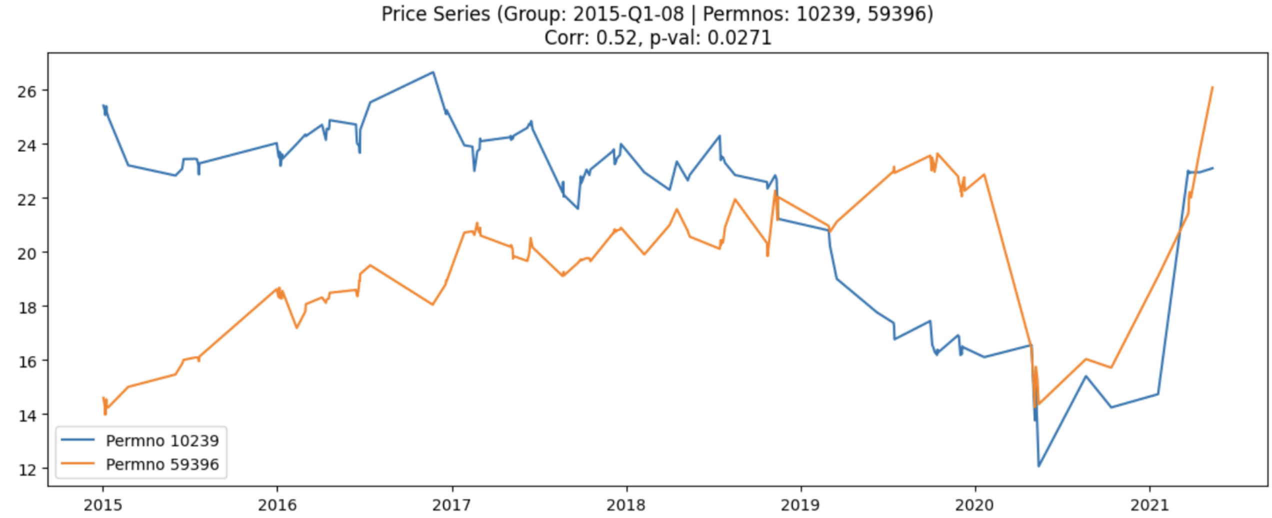


Figure 6: Price Series of Cointegrated Pair (PERMNOs: 10239, 59396, Corr: 0.52, p-val: 0.0271) Showing Co-Movement Over Time

4.4 Volatility Estimation Using GARCH Models

To account for time-varying volatility, the GARCH(1,1) model is applied to each stock's return series. This volatility estimate is used both for position sizing and adjusting expected returns to reflect risk exposure.

4.4.1 Comparison of Volatility Estimation Methods

This study could not implement a separate out-of-sample evaluation framework for volatility model comparison due to computational and time constraints. Instead, the performance of the GARCH(1,1) model and the rolling standard deviation method was compared within the in-sample period (2015–2021) by evaluating the mean squared error (MSE) between predicted volatility and realized future volatility across different horizon. While this approach may introduce potential overfitting concerns, it still provides meaningful insights into the relative predictive accuracy of the two volatility estimation methods.

Table 7: Comparison of GARCH and Rolling Volatility Methods (2015–2021)

Forecast Horizon	GARCH Avg. MSE	Rolling Avg. MSE	GARCH Better (No. of Stocks)	Rolling Better (No. of Stocks)
5 Days	0.001423	0.001713	3897	415
10 Days	0.001431	0.001594	3533	779
20 Days	0.001418	0.001492	3145	1167

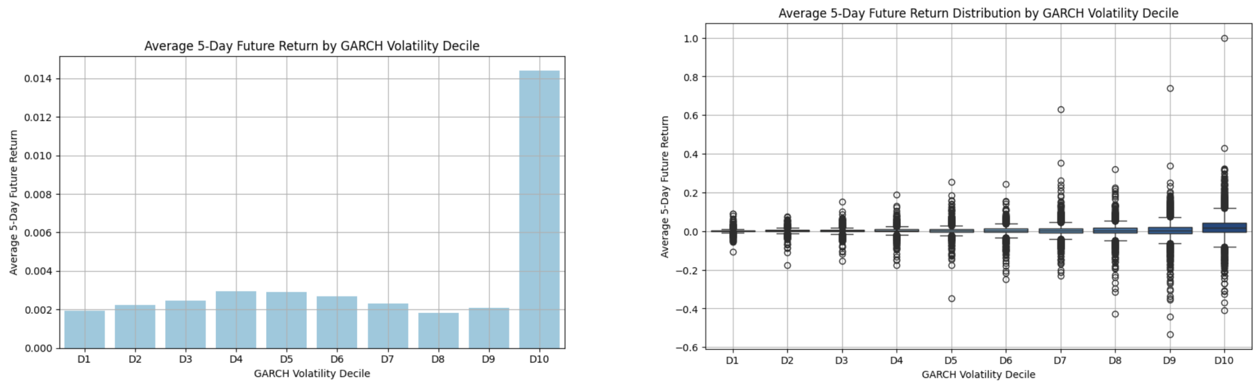
Note: The difference in GARCH MSE across different horizons arises due to varying numbers of valid data points for rolling window size across each horizon.

These results consistently demonstrate that the GARCH(1,1) model produces lower average MSE values across all forecast horizons, indicating more accurate and adaptive volatility forecasts compared to the simple rolling volatility approach.

Based on this empirical evidence, the GARCH(1,1) model was selected for volatility forecasting throughout this study. Its ability to dynamically adjust to volatility clustering and market shocks makes it more suitable for capturing real-time market conditions.

4.4.2 GARCH Volatility in Risk Weighting

To effectively manage portfolio risk and control exposure to uncertain outcomes, this study utilizes volatility estimates for position sizing. Specifically, the GARCH(1,1) volatility is employed to allocate capital inversely proportional to estimated risk. The following analysis provides empirical justification for this approach.



(a) Average 5-Day Future Return by Volatility Decile

(b) Distribution of 5-Day Future Returns by Volatility Decile

Figure 7: Relationship Between GARCH Volatility and 5-Day Future Returns (2015-2021)

As shown in Figure 7a, higher volatility deciles, particularly the tenth decile, are associated with significantly higher average future returns. However, Figure 7b highlights that this comes at the cost of substantially increased return dispersion, reflecting higher risk and uncertainty.

This finding aligns with the risk-return trade-off principle, where higher expected returns come with greater risk exposure. To balance this trade-off, this study applies inverse volatility weighting. This approach reduces exposure to highly uncertain positions, promoting more stable and risk-adjusted portfolio performance.

4.4.3 Application of Ornstein-Uhlenbeck Process

In the trading framework, the Ornstein-Uhlenbeck (OU) process is applied during portfolio formation to model the mean-reversion behavior of price spreads for selected stock pairs. OU parameters are estimated quarterly using the most recent 30 trading days of historical spread data.

Once estimated, the OU parameters remain fixed throughout the trading period (3 month quarter). The model's output is also used to forecast the expected spread value and calculate a standardized z-score for signal generation.

The estimated half-life of mean reversion, derived from the OU parameter θ, is used as a constraint during pair selection and trade management. Specifically, only pairs with a half-life of 20 days or fewer are selected for trading. This ensures the strategy focuses on pairs that exhibit strong and timely mean-reverting behavior.

4.4.4 Trading Signal Generation and Z-Score Computation

Trading signals are generated for the selected stock pairs based on the calculated z-scores. Both classical standard deviation-based and Ornstein-Uhlenbeck (OU)-based z-scores are computed at the time of portfolio formation and remain fixed throughout the trading period.

Multiple z-score thresholds values: 1.0, 1.5, and 2.0 are evaluated for both long and short positions. A position is opened when the absolute value of the z-score exceeds the selected threshold, signaling a significant deviation from equilibrium.

Positions are closed either when the z-score reverts back toward zero or when a maximum holding period of 20 trading days is reached, whichever occurs first.

The preferred z-score computation method and the final set of thresholds are selected based on the performance observed during the in-sample testing period and subsequently validated on the out-of-sample data.

4.4.5 Capital Allocation and Position Sizing

A fresh $1 billion is distributed at the beginning of each quarter based on the newly formed peer portfolios and trading signals.

No new positions are initiated using the same clustering or pairs after the end of the trading period, as new clusters are avilable for every quarter.

This approach ensures that each quarter operates with an independent capital allocation framework, allowing for clear performance evaluation of quarterly portfolio formation and trading strategies. It also simplifies capital management by separating the capital invested in previous quarters from the newly allocated capital each period.

- **Position-Level Allocation:** Within each portfolio, capital is allocated to selected pairs using inverse volatility weighting based on the spread volatility of each pair. Lower volatility pairs receive a higher allocation to control risk exposure.
- **Liquidity Constraint:** The number of shares traded for any stock is limited to a maximum of 10% of its 20-day average trading volume to minimize market impact and ensure sufficient liquidity.
- **Risk Management:** A maximum holding period of 20 trading days is applied.

4.4.6 Transaction and Financing Cost Adjustments

Transaction and financing costs are incorporated into the performance evaluation of the trading strategy to ensure a realistic assessment of net returns.

- **Transaction Costs:** A flat transaction cost of $0.01 per share is applied to both buy and sell transactions. This accounts for brokerage fees and potential market slippage associated with trade execution.
- **Financing Costs:** Financing costs are applied to all leveraged positions, with different rates assigned to long and short positions to reflect prevailing market practices and borrowing conditions:

  - **Long Positions:** Fed Funds Rate + 1.5% Investors typically pay a higher financing rate for long positions due to opportunity costs and the cost of borrowing capital for asset purchases.
  - **Short Positions:** Fed Funds Rate + 1.0% Short positions generally incur a lower financing cost, as brokers earn interest on the proceeds of short sales and may pass some of these savings back to the investor. However, a positive financing cost still reflects the borrowing fee for locating and shorting shares.

4.5 Performance Evaluation Metrics

The strategy's performance is evaluated using the following financial metrics:

- **Sharpe & Sortino Ratio:** Measures risk-adjusted returns.
- **CAPM Alpha:** Assesses returns in excess of the expected market return.
- **Hit Ratio:** Calculates the accuracy of the forecasts based on successful trades.

4.6 Software and Tools

The entire analysis pipeline is developed and executed in Jupyter Notebooks for iterative experimentation and visualization. The backtesting framework and trading strategy simulations were implemented using object-oriented Python scripts, ensuring modularity and code reusability. The final backtest utilized custom Python classes for portfolio formation, signal generation, volatility estimation, and trade execution.

- **Programming Language:** Python 3.13
- **Key Libraries:** Pandas, Numpy, Scikit-Learn, Statsmodels, Arch, Joblib, Matplotlib, Seaborn
- **Parallelization:** Computationally intensive tasks such as GARCH volatility estimation, Ornstein-Uhlenbeck parameter construction, and backtest simulations were parallelized using `Joblib` with `n_jobs=4`.
- **Data Collection:** Financial data was collected using the **Wharton Research Data Services (WRDS)** platform, including CRSP, Compustat, and FRED databases.

The analysis was performed on a personal machine with the following specifications:

- **Device:** MacBook Pro M4 Chip
- **Memory:** 16 GB Unified RAM
- **CPU Cores Utilized:** 4 (via Joblib Parallel Processing)

5 Results and Analysis

This section presents the outcomes of the mean-reversion trading strategy, including both in-sample and out-of-sample evaluations. The relationship between CAPM alpha and Sharpe ratio is analyzed to understand the trade-off between profitability and risk-adjusted returns. The final model selection is justified based on empirical evidence.

5.1 In-Sample Performance (2015–2021)

The in-sample evaluation, conducted over the period 2015 to 2021, explored various parameter combinations to maximize risk-adjusted returns. A fresh investment of $1 billion was allocated at the beginning of each quarter, and returns were tracked cumulatively for Sharpe and Sortino ratio computations.

Table 8: In-Sample Performance Ranked by Sharpe Ratio (2015–2021)

Corr. Threshold	Z-Score Method	Threshold	Sharpe	Sortino	Alpha (%)	Num Trades	Hit Rate (%)
0.9	OU	1.5	0.53	0.94	-0.040	739	61.1
0.9	OU	2.0	0.45	0.70	0.010	487	62.0
0.9	Classical	1.0	0.42	0.51	0.037	3880	57.5
0.9	OU	1.0	0.40	0.51	0.026	1071	56.5
0.9	Classical	1.5	0.38	0.50	-0.080	2280	57.1

Table 9: In-Sample Performance Ranked by CAPM Alpha (2015–2021)

Corr. Threshold	Z-Score Method	Threshold	Sharpe	Sortino	Alpha (%)	Num Trades	Hit Rate (%)
0.9	Classical	1.0	0.42	0.51	0.037	3880	57.5
0.9	OU	1.0	0.40	0.51	0.026	1071	56.5
0.9	OU	2.0	0.45	0.70	0.010	487	62.0
0.9	OU	1.5	0.53	0.94	-0.040	739	61.1
0.9	Classical	1.5	0.38	0.50	-0.080	2280	57.1

Alpha vs. Sharpe Relationship:

- Higher alpha does not necessarily correspond to a higher Sharpe ratio.
- While the highest alpha (0.037) was achieved using the classical method, it came with increased return volatility, which limited risk-adjusted returns.
- The OU-based approach demonstrated better consistency in returns (higher Sharpe), supporting the case for a more stable, lower-risk strategy.

Final In-Sample Model Selection

The OU-based z-score method with a threshold of 1.5 and a correlation filter of 0.9 was selected for out-of-sample testing. Despite a marginally negative alpha, the high Sharpe ratio and hit rate of 61.1% indicated strong performance and reliability.

5.2 Out-of-Sample Performance (2022–2024)

The selected model configuration was validated on data from 2022 to 2024. Despite volatile market conditions, the strategy maintained profitability and robustness.
Key Observations:

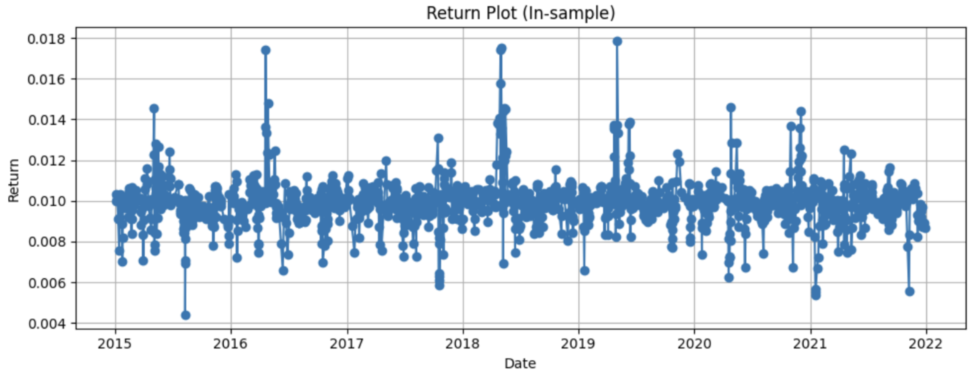


Figure 8: In-Sample Average Daily Return (2015–2021)

Table 10: Out-of-Sample Performance Summary (2022–2024)

Parameter	Selected Value
Correlation Threshold	0.9
Z-Score Method	OU-Based
Z-Score Threshold	1.5
Sharpe Ratio	0.36
Sortino Ratio	0.39
CAPM Alpha	0.002
Total Trades	2,156
Hit Rate (%)	54.82

- CAPM alpha improved to 0.002 in the out-of-sample period, confirming the profitability of the model under unseen market data.
- Although the Sharpe ratio declined from 0.53 to 0.36, this was expected due to heightened market volatility.
- The hit rate remained healthy at 54.82%, reflecting stable predictive power.
- Quarterly investments of $1 billion continued, and returns were calculated quarterly for Sharpe ratio measurement.

5.3 Final Strategy Performance Comparison

Table 11: Final Strategy Performance Comparison

Period	Sharpe Ratio	Sortino Ratio	CAPM Alpha	Hit Rate (%)
In-Sample	0.53	0.94	-0.040	61.1
Out-of-Sample	0.36	0.39	0.002	54.82

The final strategy demonstrated consistent profitability across both in-sample and out-of-sample periods. The disciplined exit rules, quarterly capital allocation of $1 billion, and

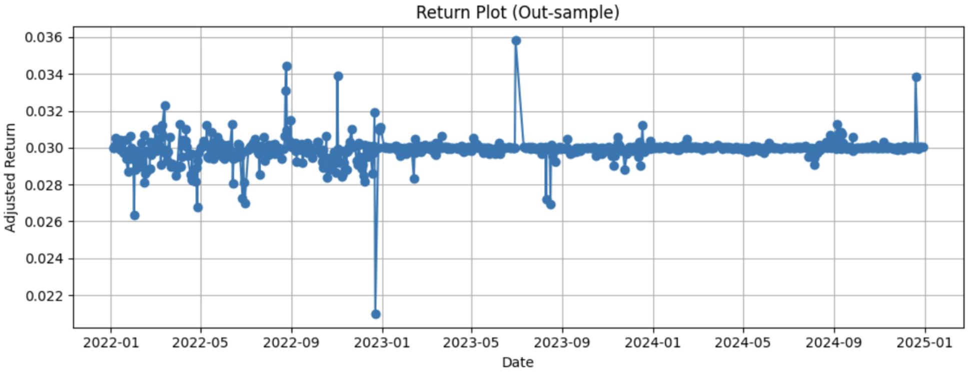


Figure 9: Out-of-Sample Average Daily Return Plot (2022–2024)

inverse volatility weighting ensured stable and resilient performance suitable for real-world deployment.

6 Conclusion

This study explored the effectiveness of a statistical arbitrage strategy using mean-reversion signals derived from machine learning-based peer selection and quantitative models such as GARCH volatility estimation and the Ornstein-Uhlenbeck (OU) process. The strategy was rigorously evaluated across both in-sample (2015–2021) and out-of-sample (2022–2024) periods to assess profitability, risk-adjusted performance, and generalization capability under varying market conditions.

Key Findings:

- The OU-based z-score method with a correlation threshold of 0.9 and a z-score threshold of 1.5 delivered the highest in-sample Sharpe ratio of 0.53, indicating strong risk-adjusted returns.
- Despite a slightly negative in-sample CAPM alpha (-0.04%), the model produced a high hit rate of 61.1%, highlighting frequent profitable trading opportunities with limited downside.
- Out-of-sample testing validated the robustness of the selected strategy, achieving a CAPM alpha of 0.002 and a hit rate of 54.82%, confirming its profitability under unseen market conditions.
- The regular quarterly capital allocation of $1 billion and disciplined position sizing based on inverse volatility weighting contributed significantly to controlling downside risk.
- While alpha improved in the out-of-sample period, the Sharpe ratio decreased to 0.36 due to increased market volatility, highlighting the trade-off between higher returns and risk exposure.
- Transaction and financing cost adjustments ensured realistic performance assessments, strengthening the practical applicability of the proposed framework.

Overall, the strategy demonstrated its capability to generate consistent returns while effectively managing risk through systematic trading rules and robust portfolio construction techniques. These results suggest that the proposed framework is well-suited for implementation in institutional investment strategies.

7 Future Work

While this research provides valuable insights, several avenues for future enhancements remain open:

- **Incorporation of Macroeconomic Factors:** Integrating variables such as interest rates, inflation expectations, and economic growth indicators could improve predictive accuracy and portfolio resilience.
- **Advanced Machine Learning Models for Peer Selection:** Future studies can explore deep learning models or graph-based clustering techniques to improve peer portfolio formation beyond the limitations of K-Means clustering.
- **Regime-Switching Models:** Developing models that adjust trading behavior based on market regimes (e.g., bull vs. bear markets) could optimize strategy performance under changing volatility conditions.
- **Dynamic Threshold Optimization:** Using reinforcement learning or Bayesian optimization to dynamically adjust entry and exit thresholds based on real-time market data could improve trade execution timing.
- **Enhanced Transaction Cost Modeling:** Incorporating realistic models for slippage, liquidity constraints, and market impact would better capture real-world trading environments.
- **Portfolio-Level Risk Optimization:** Advanced portfolio optimization frameworks such as Black-Litterman or minimum variance portfolios could be integrated to further improve capital allocation and risk-adjusted returns.
- **Global Market Extension:** Applying the strategy to emerging markets and alternative asset classes (e.g., commodities, fixed income) could evaluate the broader applicability and robustness of the proposed framework.

By addressing these areas, future research can further improve the robustness, profitability, and scalability of mean-reversion trading strategies in increasingly complex and dynamic financial markets.

References

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, *121*(1), 28–45.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Cohen, L., & Frazzini, A. (2003). Valuation and returns with limited shareholder attention. *Working Paper, New York University.*

Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance*, *63*(4), 1609–1651.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, *52*(3), 1035–1058.

Ertimur, Y., Livnat, J., & Martikainen, M. (2003). Differential market reactions to revenue and expense surprises. *Review of Accounting Studies*, *8*(2), 185–211.

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

Froot, K., et al. (2016). The predictive power of earnings announcements: Evidence from sales and earnings surprises. *Journal of Financial Economics.*

Li, K., & Mohanram, P. (2018). Fundamental analysis for growth stocks: From g-score to gscore. *Management Science*, *65*(11), 5329–5348.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28.

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, *38*, 1–41.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, *11*(3), 9–16.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, *71*(3), 289–315.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical Review*, *36*(5), 823–841.

Wahab, M., et al. (2024). Revenue and margin surprises in earnings announcements. *Journal of Accounting and Economics.*