TOBB ETÜ

University of Economics & Technology

# Arbitrage Strategy with Pair Trading: A Practical Approach to Market Neutral Trading

Project Report


**Project Team:**

- **Team Members:**
  - Pelin Buçukoğlu (191101059)
  - Burak Kutlu (201101073)
- **Department:** Computer Engineering
- **University:** TOBB ETÜ

- **Date:** 18.03.2025

## 1. Introduction

Pair trading is a statistical arbitrage strategy that identifies two assets with a historically strong relationship and exploits deviations in their relative prices. The fundamental idea is to go long on the undervalued asset while shorting the overvalued asset, assuming that their prices will revert to the mean. This strategy is market-neutral, meaning it aims to profit regardless of overall market direction.

In this project, we apply the pair trading strategy to the foreign exchange (Forex) market, utilizing historical exchange rate data to identify suitable currency pairs. Initially, **AUD/USD** and **CAD/USD** have been chosen due to their strong historical correlation. However, our approach remains flexible, allowing adjustments based on statistical analysis and market conditions. In addition to currency pairs, we also analyze several stock pairs to compare their performance. The selected stock pairs include **Apple (AAPL) & Microsoft (MSFT), SPY & QQQ (ETF indices), Coca-Cola (KO) & Pepsi (PEP), Visa (V) & Mastercard (MA),** and **Union Pacific (UNP) & CSX (**railroad companies**).** By analyzing both currency and stock pairs, we aim to determine the broader applicability of the pair trading strategy.

## 2. Objectives

The primary objectives of this project are:

- **Identify and evaluate potential currency and stock pairs based on historical relationships:** Start with stock pairs such as AAPL/MSFT and KO/PEP and currency pairs like AUD/USD and CAD/USD but remain open to better alternatives based on analysis
- **Perform correlation analysis to assess short-term relationships:** Evaluate the historical relationship between selected asset pairs in short term.
- **Test for Cointegration:** Use statistical methods such as the Engle-Granger and Johansen tests to confirm long-term relationships between pairs.
- **Develop an Algorithmic Trading Strategy:** Implement a rule-based system for executing trades automatically.
- **Establish Entry and Exit Points:** Use indicators like the z-score of the spread to determine optimal trade execution points.
- **Backtest and Evaluate Performance:** Simulate past trades using historical data to assess effectiveness.

# 3. Methodology

## 3.1 Data Collection

Historical exchange rate data is collected from Yahoo Finance, using Python's yfinance library. The dataset consists of **closing prices** for selected currency pairs, including:

- **AUD/USD** (Australian Dollar / US Dollar)
- **CAD/USD** (Canadian Dollar / US Dollar)

Additionally, stock market pairs are included for comparison:

- **Apple (AAPL) & Microsoft (MSFT)** - Technology sector.
- **SPY & QQQ** - ETF indices representing the S&P 500 and NASDAQ.
- **Coca-Cola (KO) & Pepsi (PEP)** - Beverage industry competitors.
- **Visa (V) & Mastercard (MA)** - Payment processing sector.
- **Union Pacific (UNP) & CSX** - Railroad sector companies.

The dataset spans at least 5-10 years and was cleaned by handling missing values using interpolation techniques. The historical price trends of the selected pairs are analyzed to understand their movement patterns over time (see Appendix, Figure 1).

## 3.2 Exploratory Data Analysis (EDA)

The notebook provides a detailed **EDA** to understand market behavior. Key findings:

- The dataset contains at least **5-10 years of historical data** for each currency and stock pair.
- **Correlation analysis**: The value of +1 means there exists a perfect positive correlation between the two variables, -1 means there is a perfect negative correlation and 0 means there is no correlation.

$$\rho = \frac{COV(X,Y)}{SD(X) \times SD(Y)}$$

where;
- COV(X,Y) is the covariance between assets X and Y
- SD (X) and SD(Y) = the standard deviation of the respective variables

| Pair | Correlation |
|---|---|
| AAPL and MSFT | 0.9788 |
| SPY and QQQ | 0.9927 |
| AU and AG | 0.5684 |
| KO and PEP | 0.9834 |
| V and MA | 0.9970 |
| UNP and CSX | 0.9820 |
| AUDUSD=X and CADUSD=X | 0.8800 |

- Why is correlation alone not enough?

  High correlation might be spurious, meaning two assets could move together purely by chance. Therefore, we need a stronger statistical test—cointegration—to confirm a stable relationship.

- **Other insights:**

  Trends and seasonality were examined using visualization techniques.

  Volatility and liquidity metrics were analyzed to ensure selected pairs have sufficient market activity for trading.

## 3.3 Statistical Analysis & Cointegration Testing

To determine whether selected pairs exhibit a stable long-term relationship, we follow these steps:

- **Correlation Analysis:** Identify currency and stock pairs that have historically moved together.
- **Engle-Granger Two-Step Method:**
  1. Perform an Ordinary Least Squares (OLS) regression between the two assets in each pair.

| Pair | Cointegration | Is cointegrated |
|---|---|---|
| AAPL and MSFT | 0.1539 | No |
| SPY and QQQ | 0.3074 | No |
| AU and AG | 0.6695 | No |
| KO and PEP | 0.0044 | Yes |
| V and MA | 0.0001 | Yes |
| UNP and CSX | 0.0335 | Yes |
| AUDUSD=X and CADUSD=X | 0.1749 | No |

- **Results:**

  **V & MA** show strong cointegration.

  **UNP & CSX** also exhibit meaningful long-term relationships.

- **Compute the Spread:**
  1. The spread between the two assets in a pair is calculated to determine deviations.

2. $$Spread = Asset1 - \beta \times Asset2$$

- Asset1: Price of the first asset
- Asset2: Price of the second asset
- β (Hedge Ratio): The coefficient that defines the relationship between the two assets

3. The spread movements over time were examined to determine stationarity and mean-reverting behavior (Appendix, Figure 2).

- **ADF Test p-value (Augmented Dickey-Fuller Test):**

- **p-value < 0.05** → The spread is stationary (✅ suitable for mean-reverting strategies).

- **p-value > 0.05** → The spread is not stationary (❌ unpredictable trades)

| Pair | Hedge Ratio | Mean Spread | Standard Deviation | ADF Test |
|---|---|---|---|---|
| KO and PEP | 0.2845 | 9.9618 | 1.8316 | 0.0008 |
| V and MA | 0.5534 | 20.8729 | 5.1890 | 0.0000 |
| UNP and CSX | 5.9485 | 19.2026 | 10.5137 | 0.0083 |

## 3.4 Building the Trading Strategy

- **Calculate the Z-Score of the Spread:**
  - $$Z = \frac{(Spread - Mean(Spread))}{Std(Spread)}$$
  - The moving average and standard deviation are computed over a rolling window.
- **Define Trade Rules:**
  - **Entry:** Open a position when the z-score exceeds a threshold (e.g., ±2 standard deviations).
  - **Exit:** Close the position when the z-score returns to a neutral range (e.g., ±0.5 standard deviations).

The trade entry and exit points were identified using the z-score method (Appendix, Figure 3).

## 3.5 Backtesting & Performance Analysis

The strategy was tested using historical data, and the notebook evaluated performance using:

- **Total Profit and Loss (PnL):** Measures the overall profitability of the strategy.
- **Sharpe Ratio:** Assesses risk-adjusted returns.
- **Maximum Drawdown:** Evaluates the worst-case loss scenario.
- **Win Rate:** Calculates the percentage of successful trades.
- **Annualized Return:** Evaluates the strategy's long-term potential.

| Pair | Total PnL | Sharpe Ratio | Maximum Drawdown | Win Rate | Annualized Return |
|---|---|---|---|---|---|
| KO and PEP | 2.3307 | -0.2139 | -23.6514 | 15.26% | 0.0093 |
| V and MA | 19.4121 | -0.0310 | -60.6427 | 12.56% | 0.0798 |
| UNP and CSX | -119.5645 | -0.1651 | -309.2711 | 14.50% | 0.3850 |

A comparative analysis of trading performance across Forex and stock markets is provided in Appendix, Figure 4.

## 4. Outcomes

By implementing and testing this strategy, we expect to:

- Gain insights into the effectiveness of pair trading in Forex and stock markets.
- Identify how different market conditions impact performance.
- Assess real-world challenges such as trading costs and execution delays.
- Compare profitability between Forex and stock market pairs.
- Improve strategy efficiency through optimization.

| Pair | Final Cumulative Return | Sharpe Ratio |
|---|---|---|
| KO vs PEP | 2.3307 | 0.02 |
| V vs MA | 19.4121 | 0.10 |
| UNP vs CSX | -119.5644 | -0.14 |

**Best Performing Pair:** V vs MA, (Sharpe Ratio: 0.10)

**Worst Performing Pair:** UNP vs CSX (Sharpe Ratio: -0.14)

## 5. Findings / Conclusion

This project demonstrates the effectiveness of pair trading across Forex and stock markets. Among the tested pairs, V vs MA exhibited the best performance due to its stable spread and mean-reverting behavior, making it ideal for a pair trading strategy. In contrast, UNP vs CSX showed the worst performance, likely due to higher market volatility and non-stationary price movements. Future work could explore dynamic hedge ratio adjustments and alternative exit strategies to improve overall performance.

## 6. Changes and Improvements After the Presentation

Following feedback received during the presentation, the following modifications were implemented:

- The start and end dates were changed from 2014-01-01 to 2024-01-01 to 2014-06-01 to 2024-06-01. This modification significantly altered most test results and outputs. Extending the dataset to include 2025 would likely yield different results, as some pairs may no longer pass the cointegration test, while others could qualify. Some pairs may no longer pass the cointegration or other statistical tests, while new pairs could meet the criteria.
- Enhanced the visualization of trade entry and exit points.

# 7. Appendix

## Figures and Charts

- Figure 1: Historical exchange rate trends: This figure displays the historical price movements of the selected currency and stock pairs over the given time period. It helps visualize trends, volatility, and potential correlations between assets, which are essential for identifying cointegrated pairs.

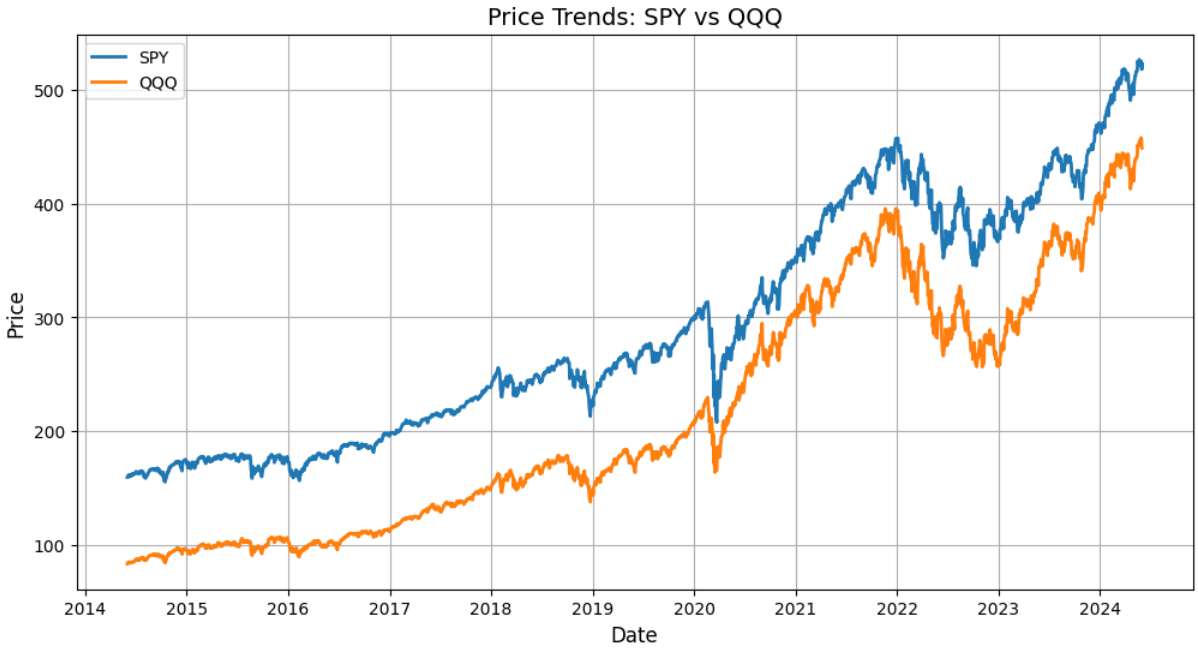


Figure 1.1: Price Trends between SPY and QQQ

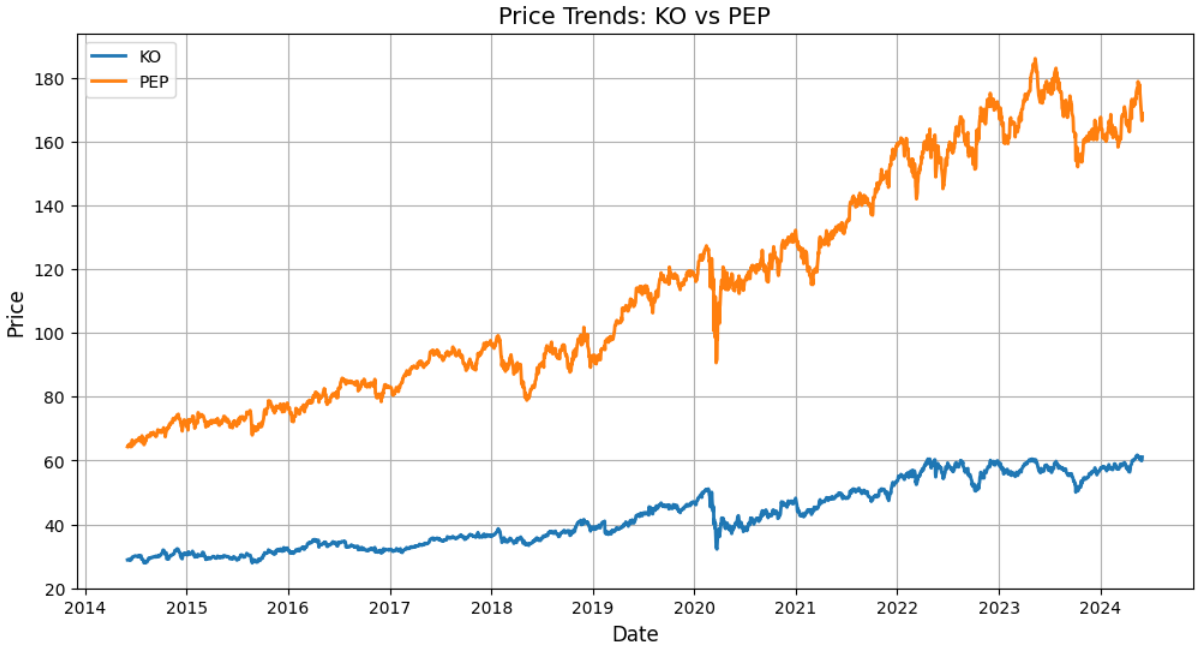


Figure 1.2: Price Trends between KO and PEP

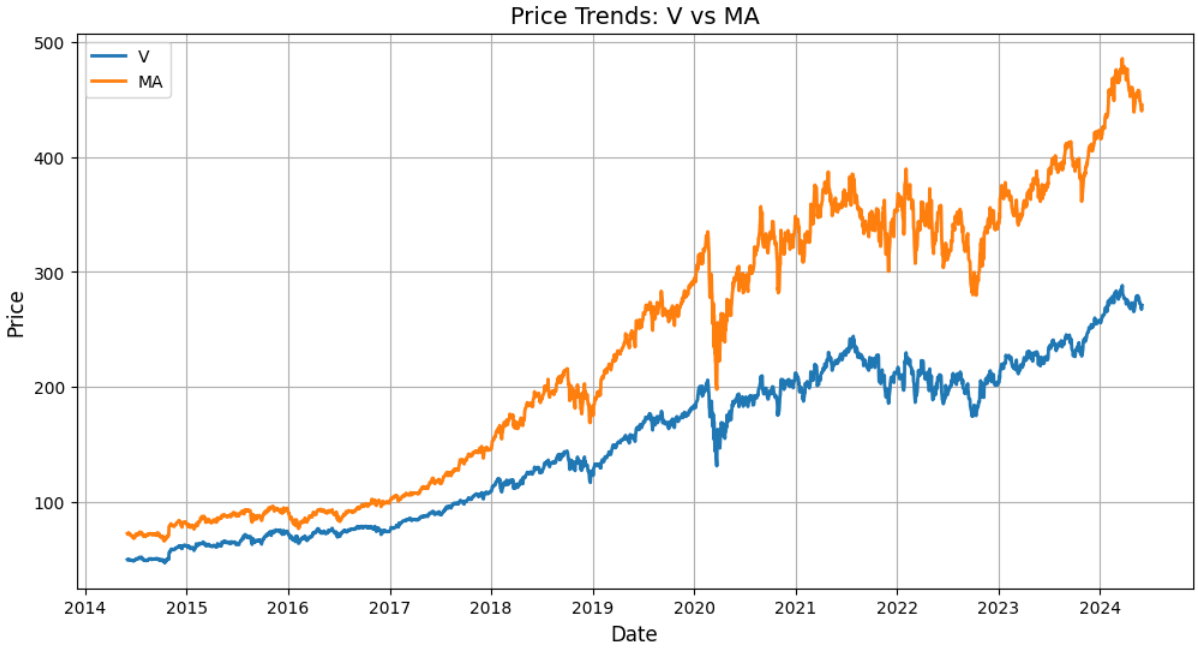


Figure 1.3: Price Trends between V and MA

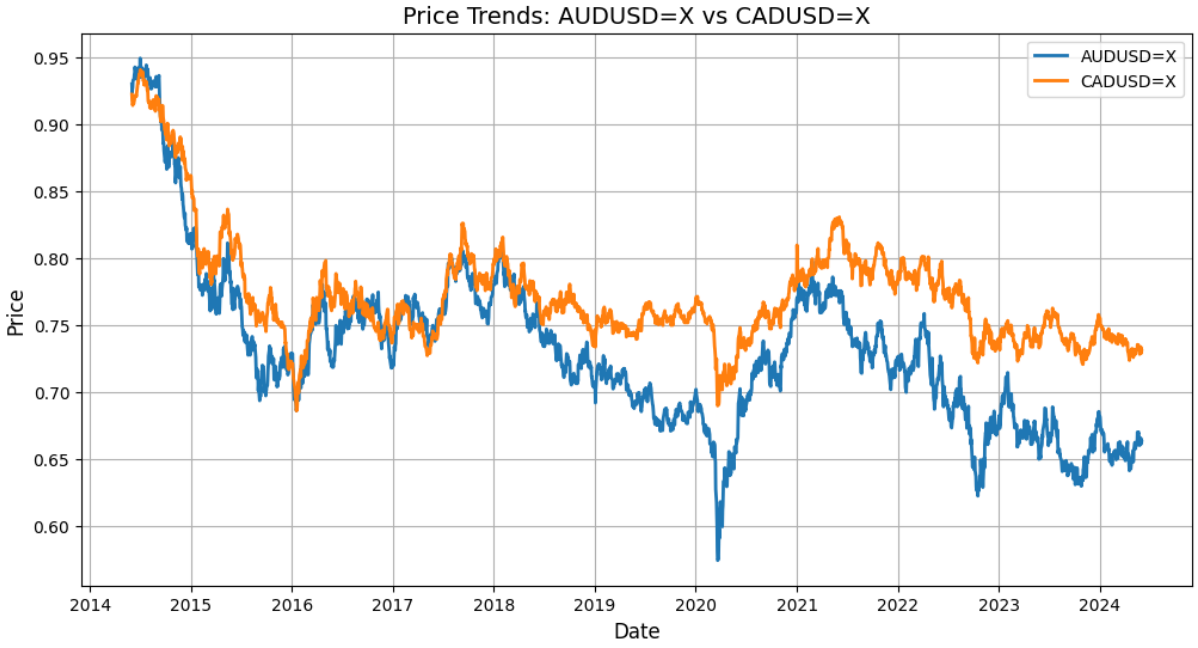


Figure 1.4: Price Trends between AUDUSD=X and CADUSD=X

- Figure 2: Spread movements over time: This figure illustrates the spread movements between selected pairs. It is a key component in identifying deviations from the mean and determining potential trading opportunities in a pair trading strategy.

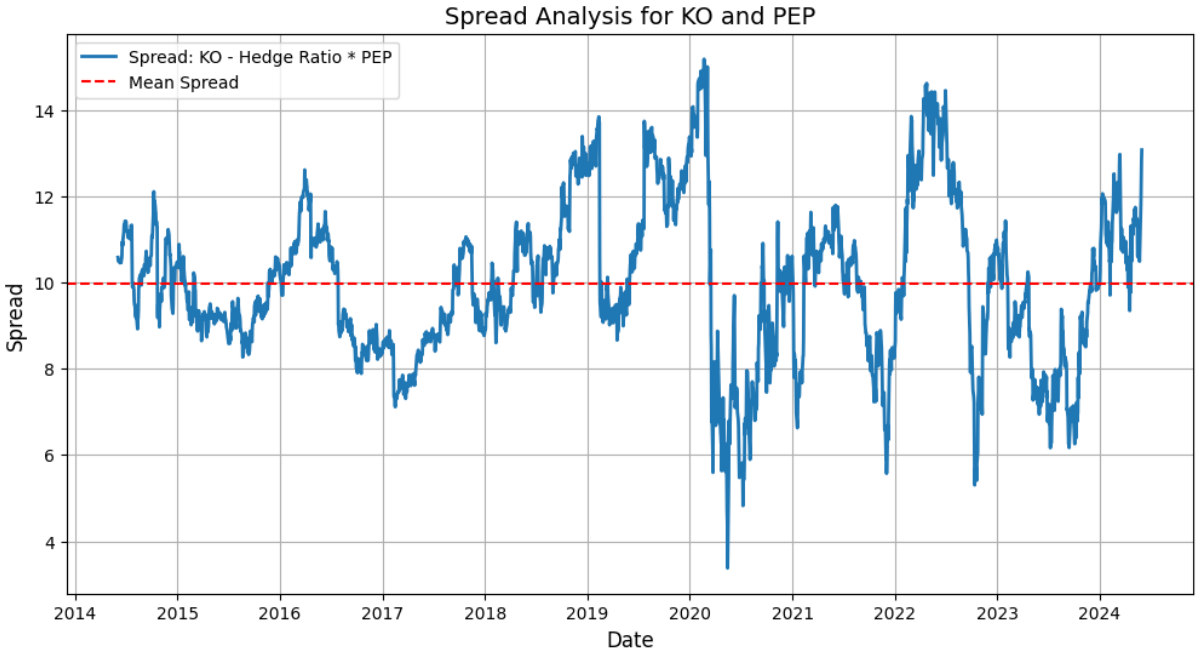


Figure 2.1: Spread analysis for KO and PEP

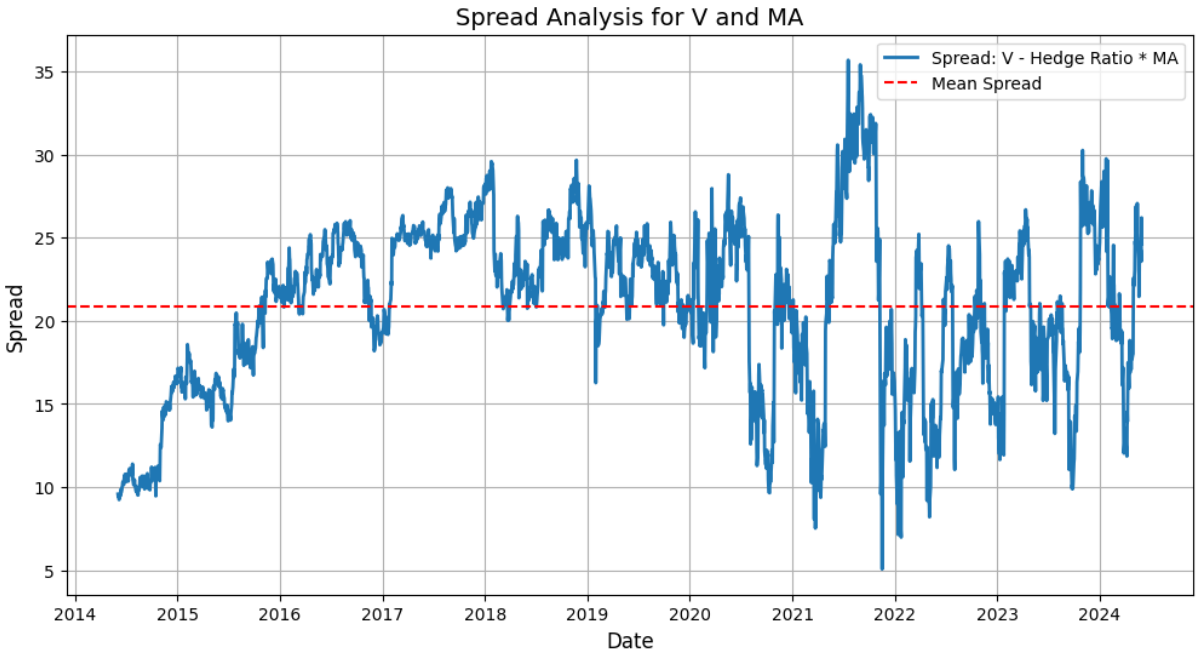


Figure 2.2: Spread analysis for V and MA

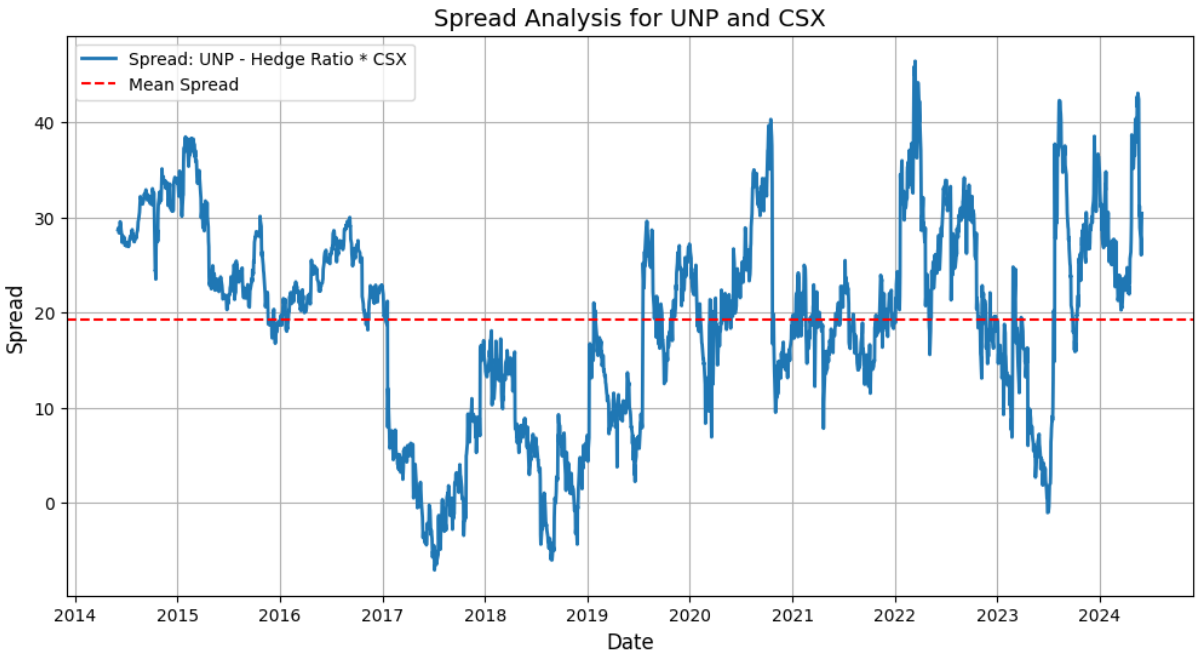


Figure 2.3: Spread analysis for UNP and CSX

- Figure 3: Trade entry and exit points: This figure highlights the buy and sell signals generated by the strategy. When the z-score exceeds predefined thresholds, trades are initiated, and positions are closed when the z-score reverts to the mean.


Figure 3.1: Z-score of KO and PEP

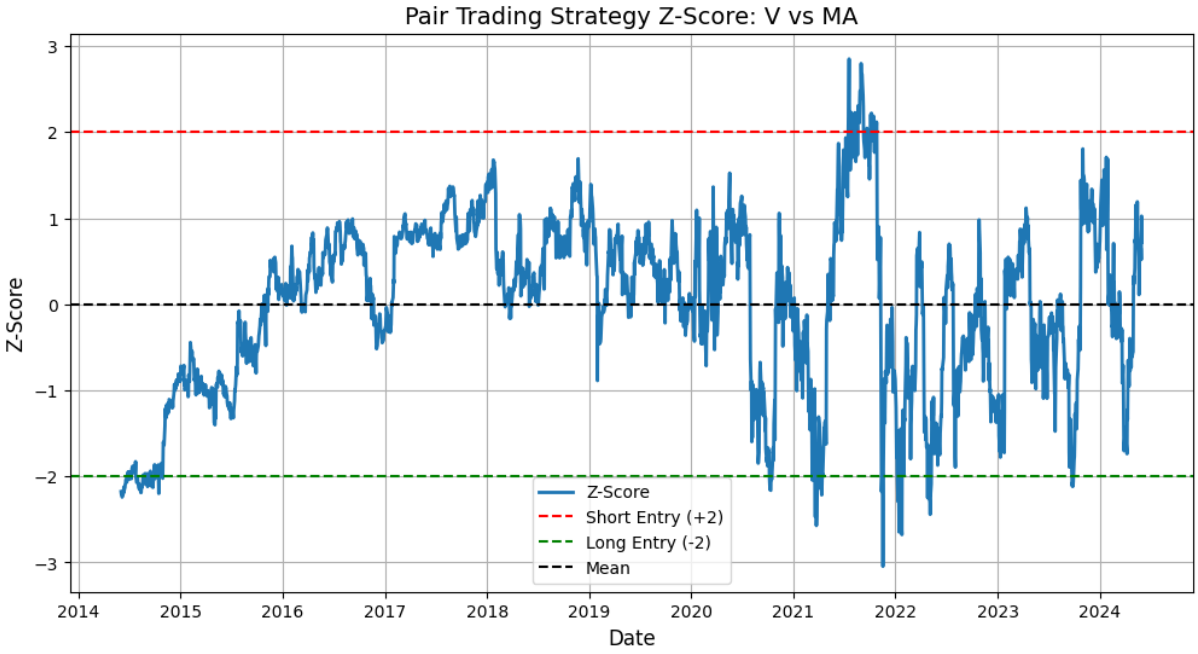


Figure 3.2: Z-score of V and MA

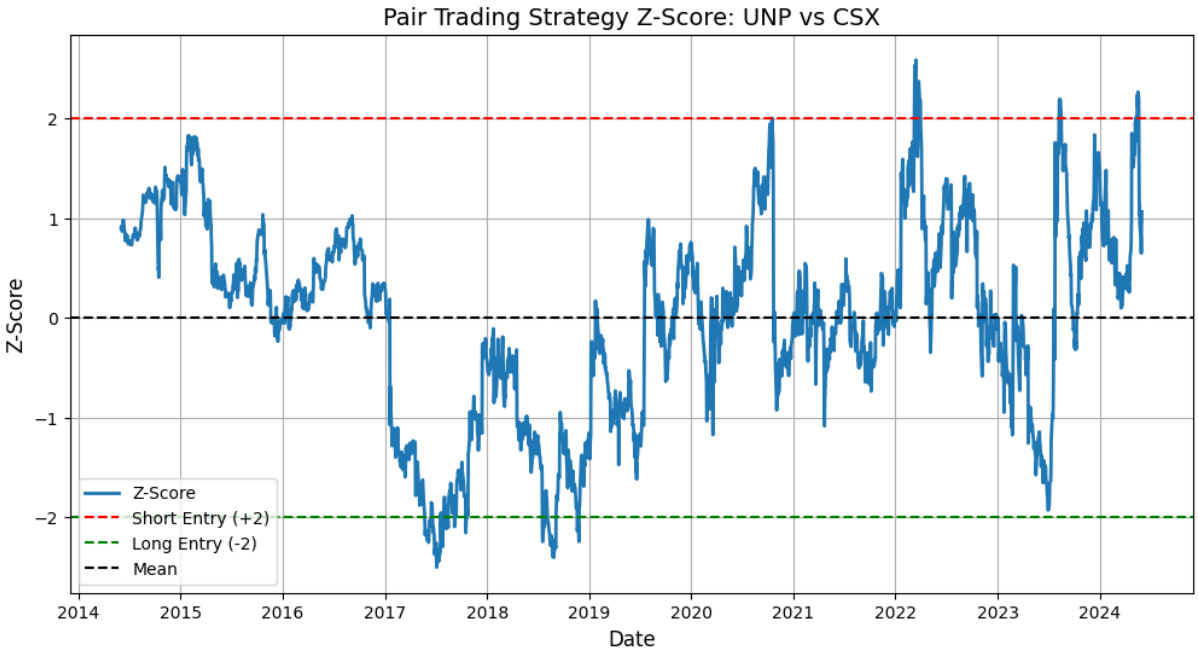


Figure 3.3: Z-score of UNP and CSX

- Figure 4: Performance comparison between Forex and stock markets: This figure presents a comparative analysis of the trading performance between Forex and stock market pairs. The metrics include total cumulative return, Sharpe ratio, and maximum drawdown for each tested pair. The comparison provides insights into which market exhibited better risk-adjusted returns and overall profitability under the pair trading strategy

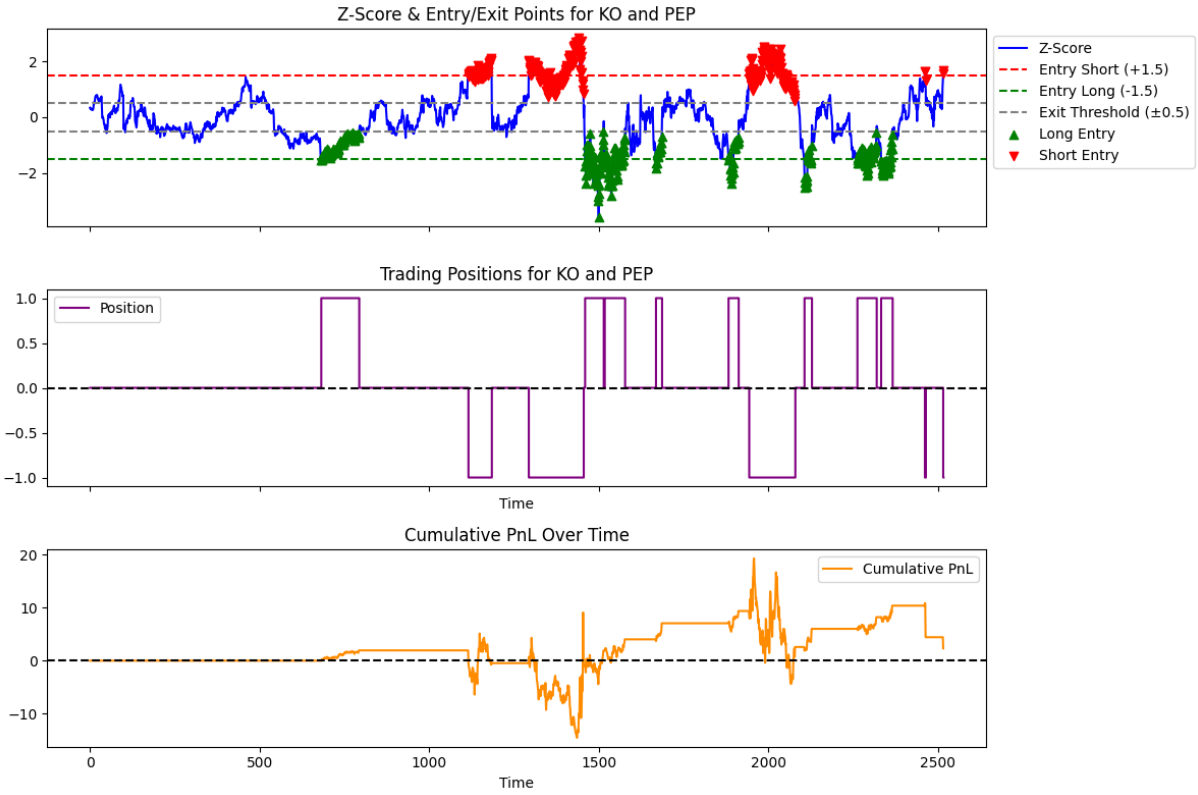


Figure 4.1: Performance of KO and PEP

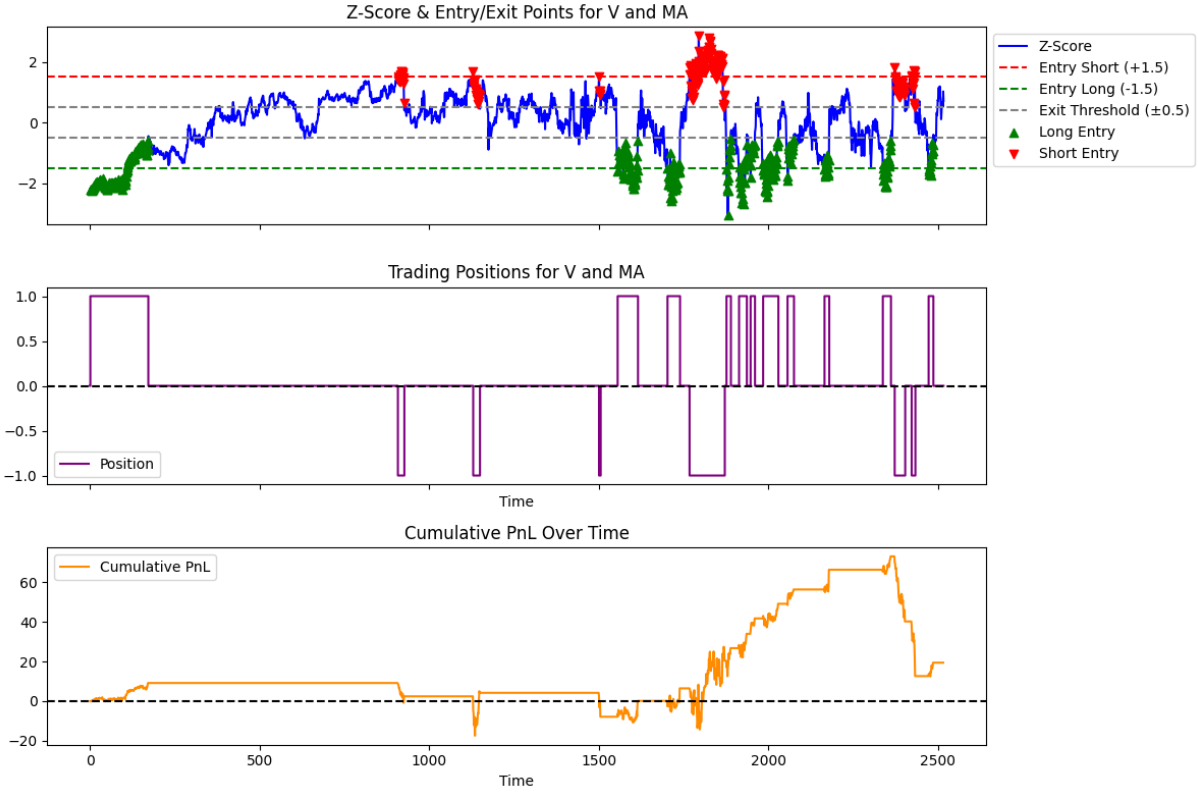


Figure 4.2: Performance of V and MA

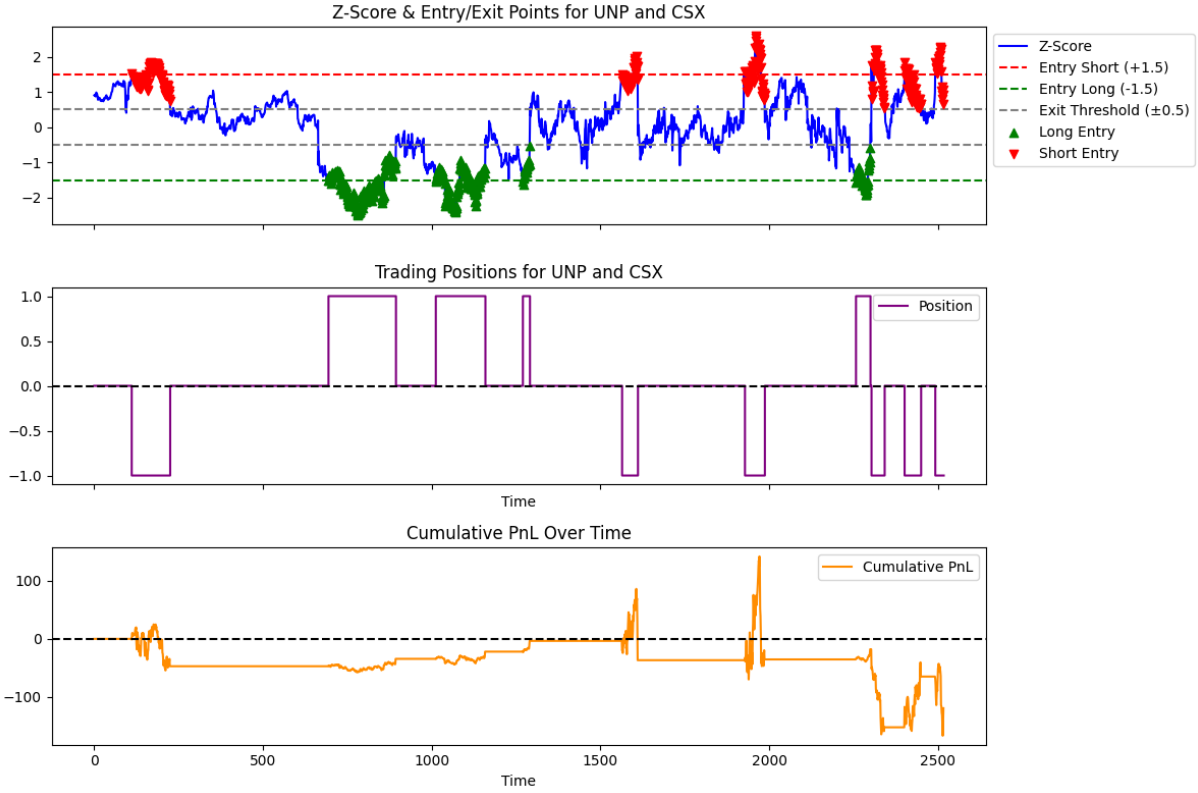


Figure 4.3: Performance of UNP and CSX