Pairs Trading Strategy Research Report

May 14, 2025

1 Implement a Basic Pairs Trading Strategy

1.1 Identify pairs of assets

I've chosen Coca-Cola (KO) and Pepsi (PEP) as my trading pair because:

1. They are in the same industry (beverages) and have similar scale
2. The business models of the two companies are highly similar
3. The historical price trend is highly correlated
4. Both are S&P 500 stocks and have good liquidity

1.2 Calculate the spread for the pairs

Data Source: Yahoo Finance API (daily adjusted close prices) The raw data is shown in the following graph 1. Note that there is a large gap in the original prices. To get a fair spread, I normalized the price data by dividing each number by the first number (times 100).

```
normalized_data = data.div(data.iloc[0]).mul(100)
```

Normalization: Base-100 normalization at inception date The price after standardization is as shown in figure 2:

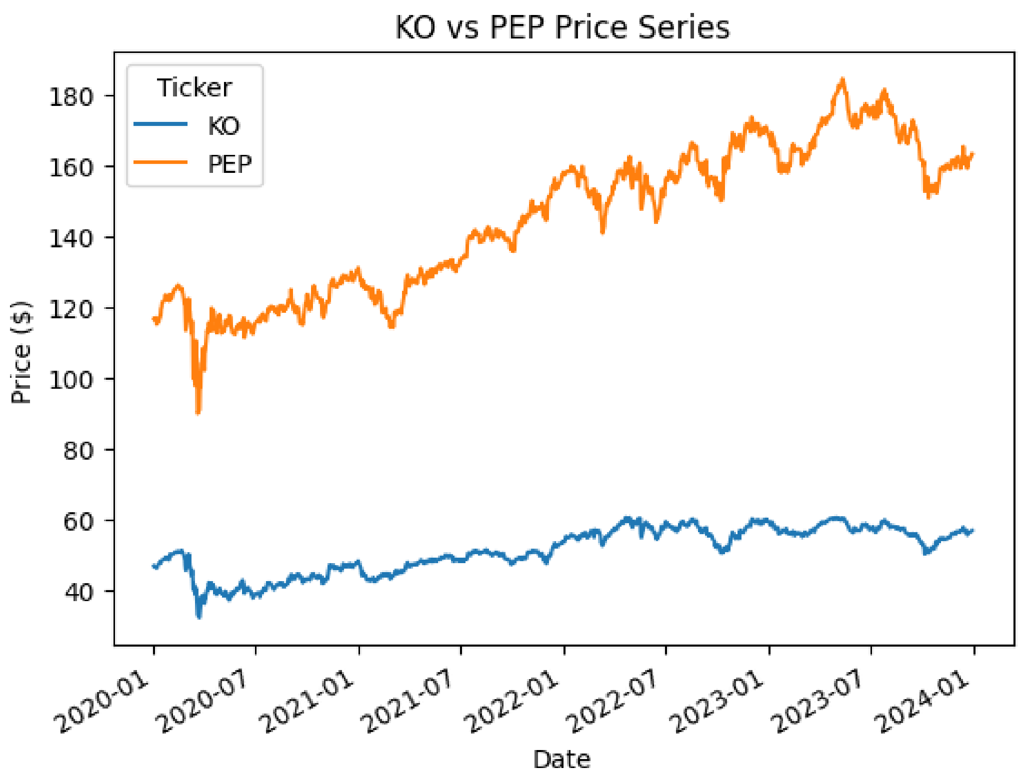

Figure 1: KO/PEP Daily Close(adj)

1.3 Define a trading rule based on the spread

The normalized spread is computed as:

$$S_t = P_{KO,t} - P_{PEP,t} \tag{1}$$

where the Price data is already normalized. Initially, the threshold was defined as 1 standard deviation from the mean. The mean and standard deviation here are calculated on a rolling basis, with a rolling window of 20 days. Note that the spread S_t can also be defined as its opposite, since we consider the change in spread to customize the trading strategy. The following figure 3 describes the spread and the trading threshold.

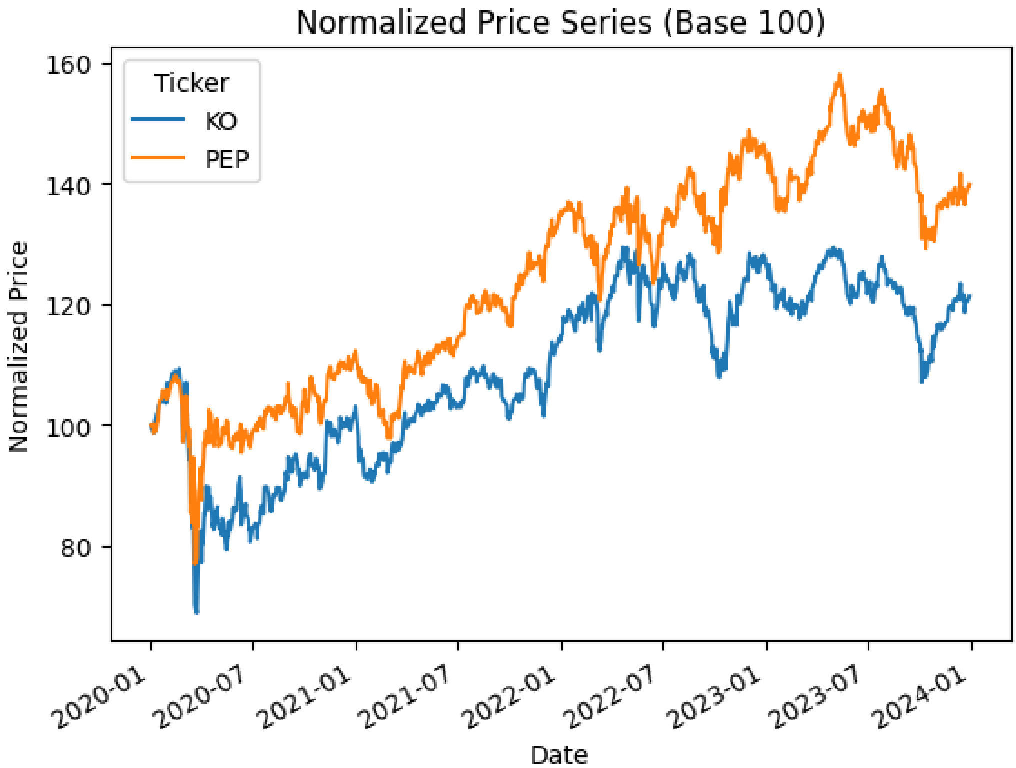

Figure 2: KO/PEP Daily normalized Close(adj)

1.4 Generate a portfolio of positions (long/short)

Trading signals follow:

$$\text{Signal}_t = \begin{cases} +1 & \text{if } S_t < \mu_t - k\sigma_t \\ -1 & \text{if } S_t > \mu_t + k\sigma_t \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where:

- μ_t: 20-day rolling mean
- σ_t: 20-day rolling standard deviation
- $k = 1$: Threshold multiplier

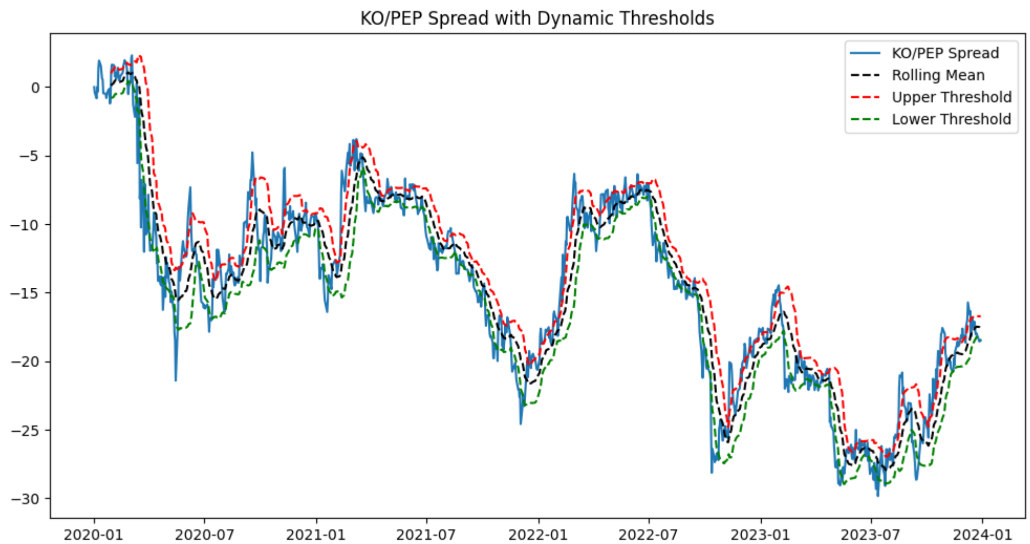

Figure 3: KO/PEP Spread

The following graph 4 illustrates the trading signals (blue: -1, short KO/long PEP red: +1, short PEP/long KO):

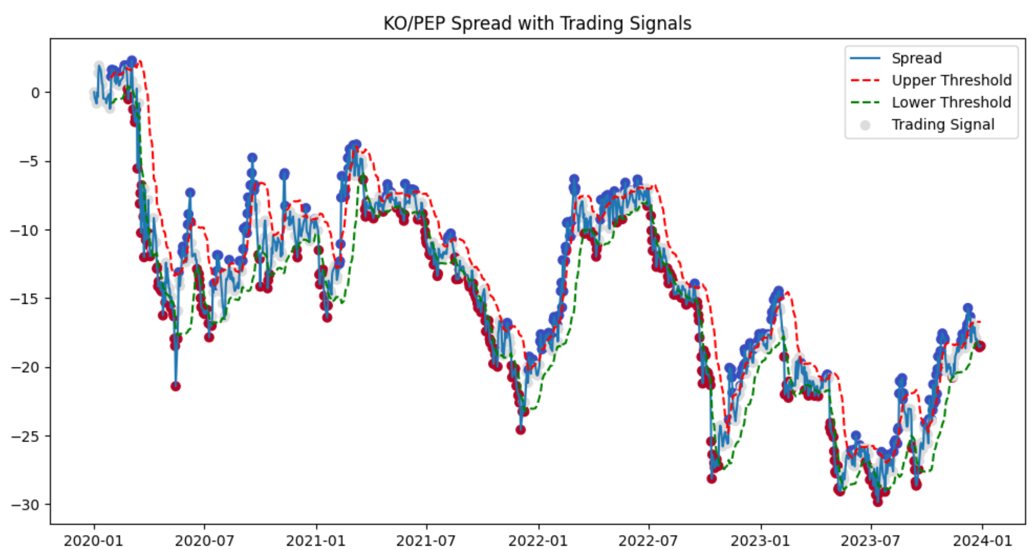

Figure 4: Trading Signal

The visualization results are shown in figure 5.

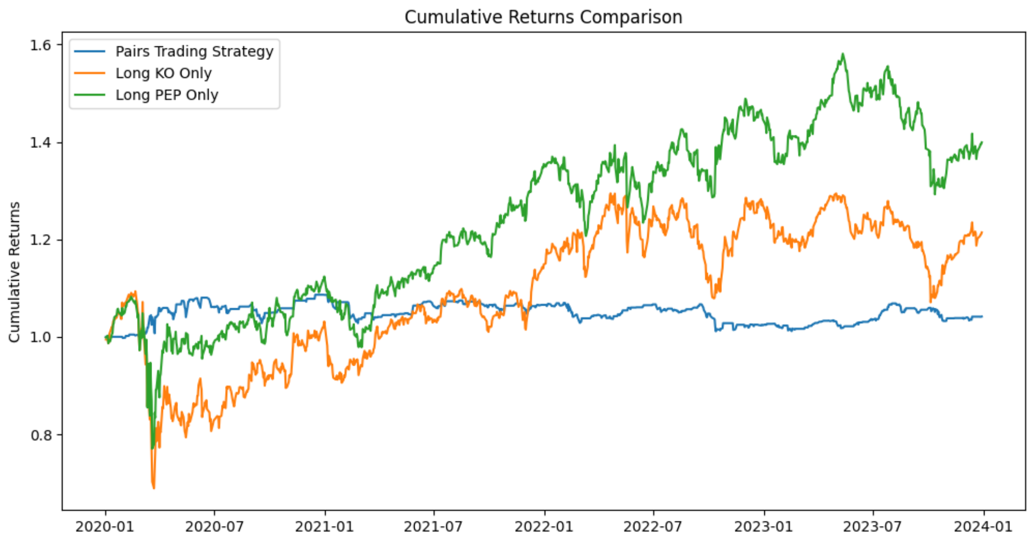

Figure 5: Return

2 Assumptions with limitations

1. There is no transaction costs for every trades. This leads to inflated returns.
2. The threshold multiplier is constant. This is not very flexible to the data.
3. Not considering the impact of slippage. This will lead to a deviation in the actual rate of return.

3 Highlight Opportunities for Improvement

1. Use Dynamic Threshold multiplier, e.g.:

$$k_t = 1.5 \times \left(1 + \frac{\sigma_t}{\mu_t}\right) \tag{3}$$

2. Add appropriate stop loss strategy.
3. Introduce slippage model.
4. Use some forecasting model to predict the spread or use the model to train some parameters.

4 Conclusion

This project implemented a Pairs trading strategy for KO and PEP stocks, analyzing their historical correlation and evaluating the strategy's performance. While the approach demonstrated potential in identifying entry and exit points, several limitations were identified, including fixed threshold assumptions (1x rolling standard deviation), the absence of transaction costs, and a lack of risk management mechanisms.

To enhance robustness, future work could incorporate dynamic threshold adjustments, transaction cost modeling, and stop-loss mechanisms. Additionally, leveraging the observed correlation between KO and PEP for a pairs trading strategy might improve risk-adjusted returns. Further validation through out-of-sample testing and sensitivity analysis on key parameters (e.g., moving average window, standard deviation multiplier) would strengthen the strategy's reliability.

Overall, the results suggest that it can be a useful tool for mean-reversion strategies, but real-world implementation requires careful consideration of costs, risk controls, and adaptive parameter optimization.

Appendix

Dependencies

- Python 3.13
- pandas, yfinance, etc.