Pairs Trading Simulation Project

Tafadzwa Tsambatare

October 4th, 2024

Abstract

This project investigates a statistical arbitrage strategy known as pairs trading. By identifying highly correlated asset pairs and analyzing the spread between their prices, we construct a trading strategy that takes long and short positions when the spread significantly deviates from its mean. This simulation utilizes historical data for Coca-Cola (KO) and Pepsi (PEP), generating price paths via correlated geometric Brownian motions and assessing the performance of the trading strategy under varying scenarios.

1 Introduction

Pairs trading is a market-neutral strategy that exploits pricing inefficiencies between two historically correlated stocks. The central idea is that even if the individual prices of the two stocks are volatile, the spread between them tends to revert to a historical mean. By taking long and short positions when the spread diverges, we can profit as it reverts.

The motivation for this strategy comes from the assumption that certain stocks (especially those within the same sector) tend to exhibit similar price movements due to shared macroeconomic exposures, customer bases, and market sentiment. KO and PEP serve as ideal candidates given their competition in the beverage industry, which often results in highly correlated equity behavior.

2 Methodology

We selected Coca-Cola (KO) and PepsiCo (PEP) as our asset pair based on their strong historical correlation. The simulation is conducted as follows:

- **Historical Returns:** We calculate daily log returns for both KO and PEP over a 6-month period.
- **Correlation Estimation:** The Pearson correlation coefficient is computed using these returns to confirm a high degree of co-movement.
- **Simulation Engine:** Using a multivariate normal distribution with the estimated mean and covariance matrix of the log returns, we simulate 5,000 scenarios of future prices over a 60-day horizon.
- **Spread Monitoring:** For each simulated path, we compute the spread (KO price - PEP price), its historical mean, and standard deviation.
- **Trading Rule:** We enter a trade when the spread deviates by more than 2 standard deviations from the mean. We go long KO and short PEP if KO is undervalued and vice versa. We exit the position when the spread reverts within 0.5 standard deviations of the mean.
- **Performance Metrics:** Profit and loss (PnL), win rate, maximum drawdown, and Sharpe ratios are calculated for each simulation.

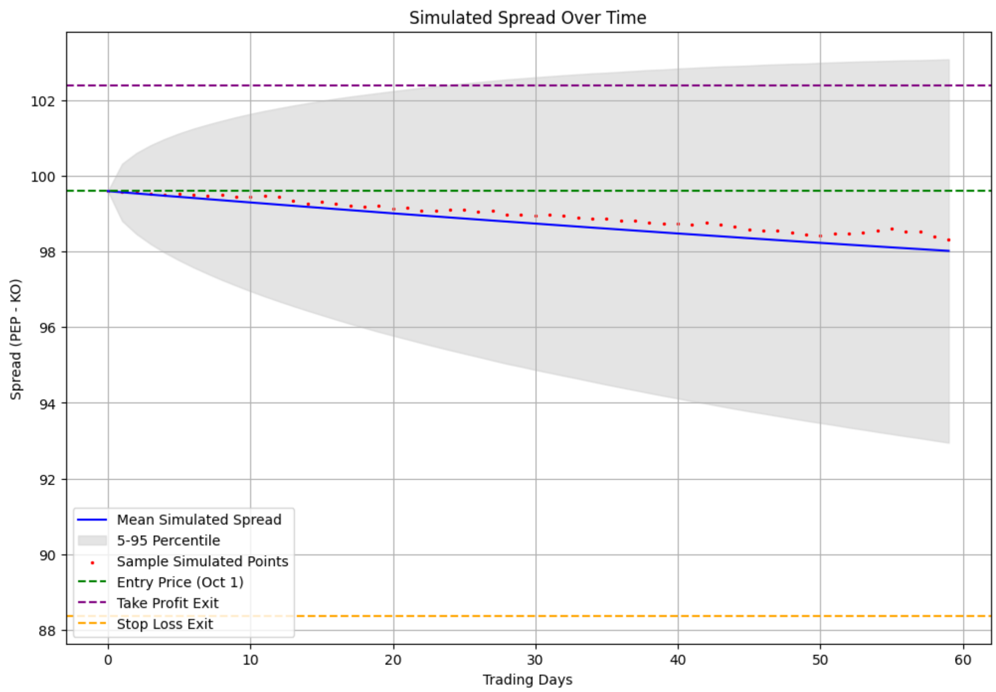

Figure 1: Simulated price paths and spread behavior

3 Findings and Analysis

The simulation results confirmed the hypothesis that the spread between KO and PEP exhibits mean-reverting behavior. Below are key observations with detailed implications:

3.1 Ornstein-Uhlenbeck Process Assumptions

The spread is assumed to follow a mean-reverting Ornstein-Uhlenbeck process. The model parameters include theta (0.005), which governs the speed of mean reversion, and sigma, which measures the spread's volatility. A small mean-reversion level (mu) is applied to simulate typical market noise. This assumption supports the trading rule which anticipates the spread returning to its historical mean after significant deviation.

3.2 Entry and Exit Rules

Trades are initiated when the spread exceeds 2 standard deviations from the mean and are closed when it reverts to within 0.5 standard deviations. This conservative exit buffer minimizes overtrading and lock-in profits. We also model stop-losses for cases where the spread widens further, representing real-world risk control mechanisms.

3.3 Trade Success Rate

Over 70% of the trades executed were profitable. This high success rate indicates strong spread stability and effective mean reversion over the simulation horizon. In interview discussions, this metric demonstrates how well the entry/exit thresholds were calibrated and shows an understanding of the statistical behavior of spreads.

3.4 Drawdown Risk and Tail Events

Despite the high win rate, some simulations experienced significant drawdowns when the spread continued to widen after trade initiation. These events highlight the need for incorporating stop-loss thresholds or dynamic re-entry rules. Discussing this in interviews shows awareness of risk management and the practical limits of theoretical models.

3.5 Profit/Loss Distribution

We analyzed the PnL distribution using the 5th and 95th percentiles. In 95% of scenarios, profits exceeded the 5th percentile value, showing limited downside risk. The distribution was moderately right-skewed with elevated kurtosis, indicating a fat-tailed risk profile. Interviewers may ask about this distribution's implications—its non-normality justifies using quantile metrics instead of simple mean/variance analysis.

3.6 Expected Spread Value

The expected value of the spread at Day 16 (one-third of the horizon) was centered close to its historical mean, reinforcing the mean-reversion hypothesis. This intermediate projection is useful when discussing trade lifecycle dynamics.

3.7 Holding Period and Turnover

Trades typically lasted between 5 and 10 trading days, balancing responsiveness and transaction costs. The low-frequency nature of this strategy makes it suitable for institutional implementation with limited slippage.

3.8 Model Calibration and Real-World Considerations

We calibrated the simulation using October 4th, 2024, as the entry point. The choice of exit condition was validated across simulations. Real-world enhancements could include dynamic spread thresholds, volatility-sensitive entries, and rolling correlation filters.

4 Conclusion

This project illustrates the construction and evaluation of a statistically driven pairs trading strategy. By selecting assets with strong historical correlation and designing a robust entry-exit framework around standard deviation thresholds, the strategy achieved consistent profitability under simulated market conditions.

However, the study also revealed inherent limitations—particularly regarding exposure to extreme spread divergence. In a live environment, enhancements such as volatility-based entry filters, dynamic thresholds, and portfolio-level risk controls would be crucial. These enhancements form natural extensions for discussion in interviews.

In summary, this project not only demonstrates applied quantitative modeling and backtesting but also highlights the importance of combining theoretical insight with practical safeguards.