Pairs Trading: Toyota and Subaru

Victoria Uriostegui and Erika Dickson,
CSUF | Computer Science Department of Computer Science, California State University, Fullerton

CS Showcase | April 8th, 2025

Background

We seek to explore pairs trading with correlated and cointegrated stocks to determine if an effective pairs trading strategy will yield profits. Pairs trading is a market neutral trading strategy that lies in the intersection of statistics, finance, and computer science[1]. If utilized correctly, pairs trading has the potential to greatly minimize loss in a volatile market, due to its ability to capitalize on both "dips" and "rips"[1]. It is due to this versatile nature that we were first drawn to it and are now researching the possibilities and limitations of pairs trading, and how far the topic can be pushed in the interests of our computer science backgrounds. In delving into the implementation of pairs trading, we use data science in tandem with statistical and financial approaches to discover the degree to which pairs trading can be computerized, to the goal of being a "hands-off" system. The research question behind our work is: to what extent can we design our implementation of pairs trading to maximize the amount of profit gained within a specified period and minimize the amount of human intervention required to operate it? The potential implications of our research include the possibility of adapting our work to help break down financial and educational barriers to the stock market and make finance more accessible by aiding others in navigating and profiting from market volatility.

Hypothesis

The goal of our research is to implement a successful, convenient pairs trading system synthesizing statistical and financial algorithms to secure a nontrivial profit utilizing a reasonable initial investment principal. Ultimately, we seek to discover if the pairs trading strategy is feasible, effective, and profitable for the average trader.

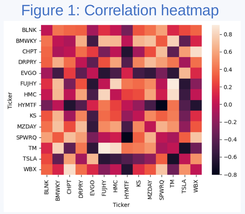
Methods

1.) Our first step includes determining our development environment; we are using the Google Colab cloud environment and a Jupyter Notebook file to run our code on due to the tools' scalability.

2.) Next, utilizing the Yahoo Finance API we pull data for a one-year period (9/20/2023 - 9/20/2024) and select stocks related to the automotive and electric charging stations company.

3.) Using Python's built in function, *corr()*, and statsmodels API, we create a correlation heat matrix (Figure 1) and find the stocks Toyota and Subaru (Figure 2) are most correlated according to their 0.9057 p-value (Figure 3).

Figure 1: Correlation heatmap



Stock 1	Stock 2	Corr Coeff
TM	FUJHY	0.905722
CHPT	WBX	0.841907
HMC	TM	0.841443

Figure 3: Correlation coefficient

Figure 2: Toyota and Subaru

4.) We implement the Engle-Granger cointegration test in two parts. Before starting with the test, we convert Toyota and Subaru stock prices to their natural logarithmic price. Next, we conduct the Ordinary Least Squares test, which utilizes the seaborn library to find the best fitted line (Figure 4). Following, using linear combination, we implement the Augmented Dicky-Fuller test and find, with a p-value of roughly 0.083, that we can confirm with minimum 90% confidence that the stocks are cointegrated because the p-value is between 0.0 and 0.10 (Figure 5).

Figure 4: Ordinary Least Squares with best fitted line

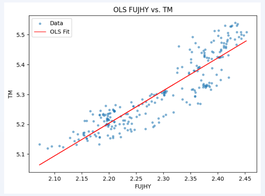


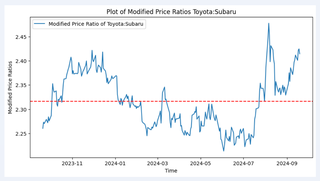
Figure 5: Augmented-Dickey Fuller p-value

```
adf_test = sm.tsa.adfuller(Zt)
print("P-value:", adf_test[1])

P-value: 0.08444085435483395
```

5.) Then, we manually calculated the z-score and natural logarithmic ratios to implement the pairs trading signal (Figure 6).

Figure 6: Toyota and Subaru natural logarithmic price ratios



6.) We then prepare the data for use in our pairs trading algorithm. The code implementation we use is largely adopted from Auquan's pairs trading GitHub repository, with adjustments to suit our project[2]. To ensure we have most relevant and recent figures, we use rolling windows of calculations from the stock price ratios data to calculate the z-score, which will serve as the pairs trading signal to determine when the system should take a long or short position on the stocks. The z-score is calculated by taking the five day mean, subtracting the 60 day mean, and dividing the difference by the 60 day standard deviation.

7.) Finally, we implement the pairs trading algorithm in our *trade_action()* function, which is also adapted from Auquan's pairs trading repository[2]. The algorithm follows real-life trade actions by taking long and short positions or clearing positions. The threshold to indicate a trade opportunity uses the z-score previously calculated. When the z-score is greater than 1, we sell short by selling the larger stock, stock one, and purchasing shares of stock two, the smaller stock, and calculating the correspond money. Similarly, when the z-score is less than 1, we buy shares of stock one and sell stock two shares and calculate the money. Clearing positions requires selling all current shares. We utilize a tracking DataFrame to record our trade actions and the ensuing monetary changes.

Results

The results of our research culminates into an overall monetary figure of $1,256.31. Considering our starting value of $1,000, this leaves us with a net profit of approximately $256.31 over a period of one year, consisting of 252 days of tracked financial data. The procedure executed for these results was the pairs trading algorithm which calculates the return on our investments and logs the changes and circumstances of each trade action into a DataFrame (Figure 7). Additionally, we track the stock prices over the course of the observed year and add an overlay of trade signals to analyze what trends contribute to trade signals (Figure 8). In terms of our goal to have minimal user interaction to run our program successfully and to fruition, users are still required to gather data and research correlated and cointegrated stocks, as well as perform their own analysis on the correlation and cointegration values. In addition, the statistical calculations for the trade signal algorithm and the trade signal algorithm itself requires adaptation to any chosen stocks that differ from those presently analyzed.

Figure 7: Trade action tracker DataFrame

	Date	Toyota	Subaru	Buy_Toyota	Buy_Subaru	Sell_Toyota	Sell_Subaru	Position	Profit Total	Profit Change
13	2023-10-09 00:00:00	5.159745	2.207175	0	1	1	0	Short	1056.556812	56.556812
14	2023-10-10 00:00:00	5.176997	2.24071	0	1	1	0	Short	1111.893331	111.893331
20	2023-10-18 00:00:00	5.17037	2.219203	0	1	1	0	Short	1168.34402	168.34402
22	2023-10-20 00:00:00	5.156062	2.182675	0	1	1	0	Short	1226.497713	226.497713

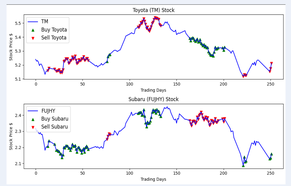


Figure 8: Tracking buy and sell signals for each respective stock

Conclusions

In conclusion, our research yields unexpected findings. Our hypothesis stipulates that our implementation of a pairs trading strategy will yield sizeable profits given two cointegrated stocks and will be convenient for the average trader to utilize. However, our findings show that although more than $200 was accrued through our pairs trading implementation, our results were not up to par with expectations and fell short of other comparable projects. Limitations encountered include using stocks that were cointegrated but had large price gaps, which made the process of purchasing and buying equivalent amounts of stocks complicated. In terms of our goal of minimizing human intervention in our program to have a relatively automated system, we fall short of our goal to simplify the process of pairs trading, as much analysis and computation must still be completed manually. Despite these setbacks, our research remains compelling in the merits of the pairs trading strategy and sets the foundation for further study.

Future Work

In future revisions, we aim to broaden the functionality and scalability of our project, as well as increase the return on investment. Implementations and edits to accomplish this goal include implementing machine learning to allow the system to learn from the pairs trading algorithm to make more sophisticated real-time decisions, adapting the program to be easier to use for the average user, and utilizing more sophisticated statistical analysis practices to have more accurate and relevant figures upon which the pairs trading strategy can be utilized, such as testing which specific rolling window expanse is most effective.

References

1. James, Chen. "Pairs Trade: Definition, How Strategy Works, and Example." Investopedia, 26 June 2021, www.investopedia.com/terms/p/pairstrade.asp.
2. Auquan. "Tutorials/Pairs Trading.ipynb at Master · Auquan/Tutorials." GitHub, 2025, github.com/Auquan/Tutorials/blob/master/Pairs%20Trading.ipynb.
3. Kramer, Leslie . "Long Position vs. Short Position: What's the Difference?" Investopedia, 2019, www.investopedia.com/ask/answers/100314/whats-difference-between-long-and-short-position-market.asp.

RESEARCH POSTER PRESENTATION TEMPLATE © 2018
www.PosterPresentations.com