Pair Trading Strategy Report

Aryan Maskara

July 2024

1 Introduction

This project implements a **Pair Trading Strategy model** for the stocks **Mastercard and Visa** and their **adjusted closing prices** over the **5-year period** July 2019 - June 2024. The data was taken from Yahoo Finance.

Name	Currency	Exchange
Mastercard Incorporated (MA)	USD	NYSE
Visa Inc. (V)	USD	NYSE

The **goal** of the project is to develop a trading strategy model that **generates profitable returns** over time.

2 Working of Model

The model uses **Pair Trading**, a **Market Neutral, Statistical Arbitrage strategy**.

2.1 Data Handling

1. Remove all dates from the datasets that are not in both datasets.
2. Remove all columns in the datasets other than the Date and Adjusted Closing Price.
3. The data is divided into a **65-35 training-testing ratio** to ensure the model works for fresh data.

2.2 Preliminary Tests

1. After the data has been divided into training and testing data, we take the **training data** and check the **correlation between the two stocks**. The correlation needs to be high enough (≥ 0.9) for us to do pair trading with the two stocks.
2. Given that the correlation is high, now we check for the **co-integration between the two stocks**. Co-integration means that even though the price of the stocks is non-stationary, a linear combination of the stock prices should be stationary. This is done by using the Augmented Dickey-Fuller test.

3. The test gives a **p-value** which denotes how likely is it for the linear combination of the stock prices to be non-stationary. A p-value of ≤ 0.1 is required to proceed with the strategy.

2.3 Model

1. We use **Ordinary Least Squares (OLS)** regression technique on the **training data** to calculate the **hedge ratio** of Mastercard compared to Visa.

2. OLS formula: Visa price = n * Mastercard price + residual. Here n is the hedge ratio that we calculated in step 1.

3. We calculate the residuals for the **training and testing data, based on the hedge ratio**. Upon normalizing the residuals, they tend to **hover around 0**. These normalized residuals is called the **z-score**.

4. We use the **z-score of the current day** to make trades and generate profits.

2.4 Market Signals

1. **Short Mastercard and long Visa**: If the z-score is **atleast 0.75 standard deviation above mean**, either Visa is underpriced or Mastercard is overpriced. Hence, we short 1 stock of Mastercard and long equivalent ratio of stocks of Visa, creating a perfectly hedged trade.

2. **Long Mastercard and short Visa**: If the z-score is **atmost 0.75 standard deviation below mean**, either Mastercard is underpriced or Visa is overpriced. Hence, we long 1 stock of Mastercard and short equivalent ratio of stocks of Visa, creating a perfectly hedged trade.

3. **Exit position**: If the z-score is **between 0.25 standard deviation of the mean**, the stocks have more or less come to a similar price as indicated by the OLS regression. Hence, we exit the position.

4. **Stop loss**: If the z-score is **above 3 on either side of the mean**, the stocks are not moving in the manner predicted. Hence, we exit the position to prevent having a huge loss.

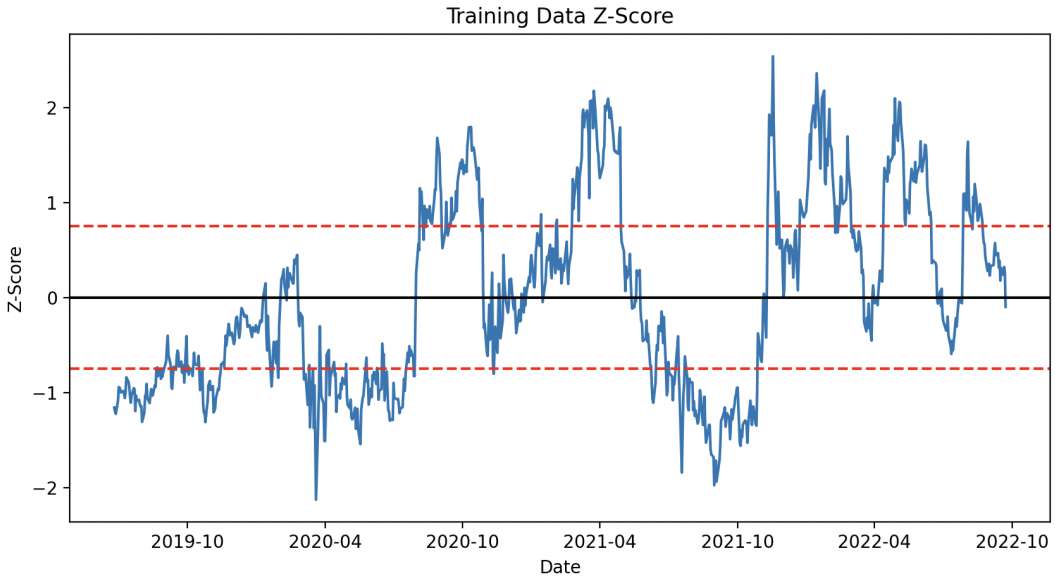
5. **Maximum capacity of stocks**: Furthermore, to ensure that there is **not a huge risk**, there is a **maximum capacity of 100 long or short Visa and Mastercard.** No trades will be made if they exceed the absolute value of 100 stocks.

3 Results

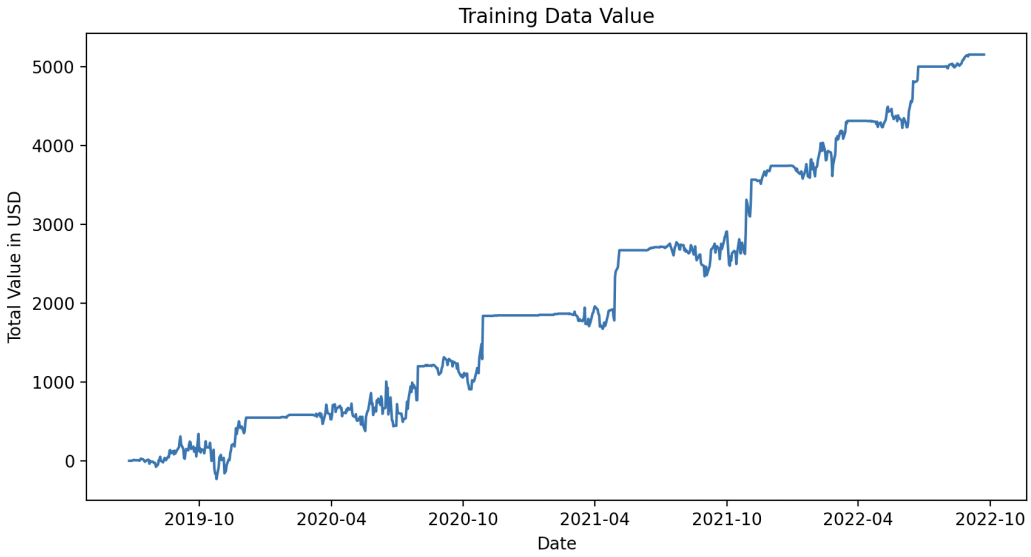
The preliminary tests yield a **correlation of 95.71%** and a **p-value of 0.093**, as shown below. This shows that the **pair of stocks are suitable for pair trading**.

```
correlation: 0.957114639133192
p-value: 0.09337879149968853
```

The following image shows the **z-score of the training data**, with the market signals to long Visa and short Visa dotted in red.

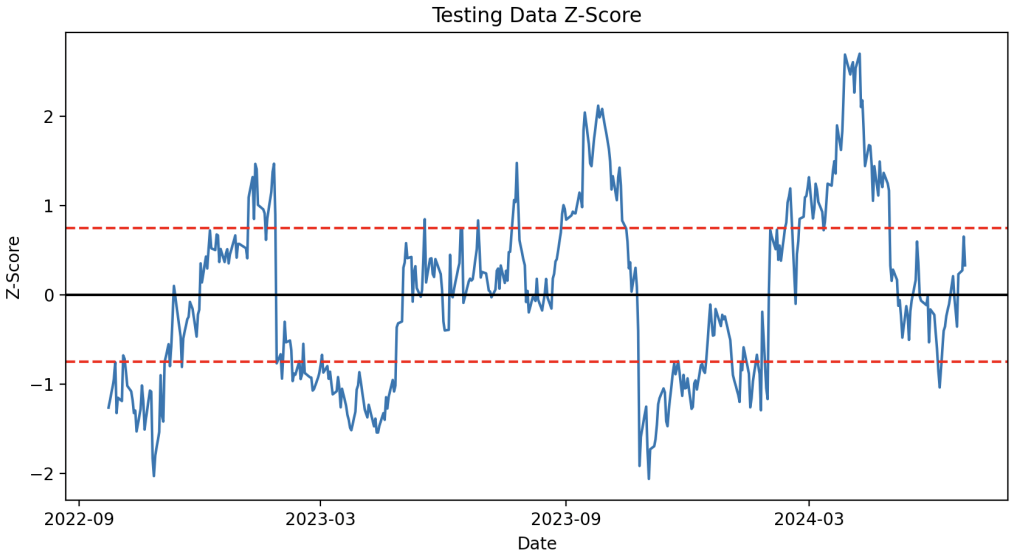


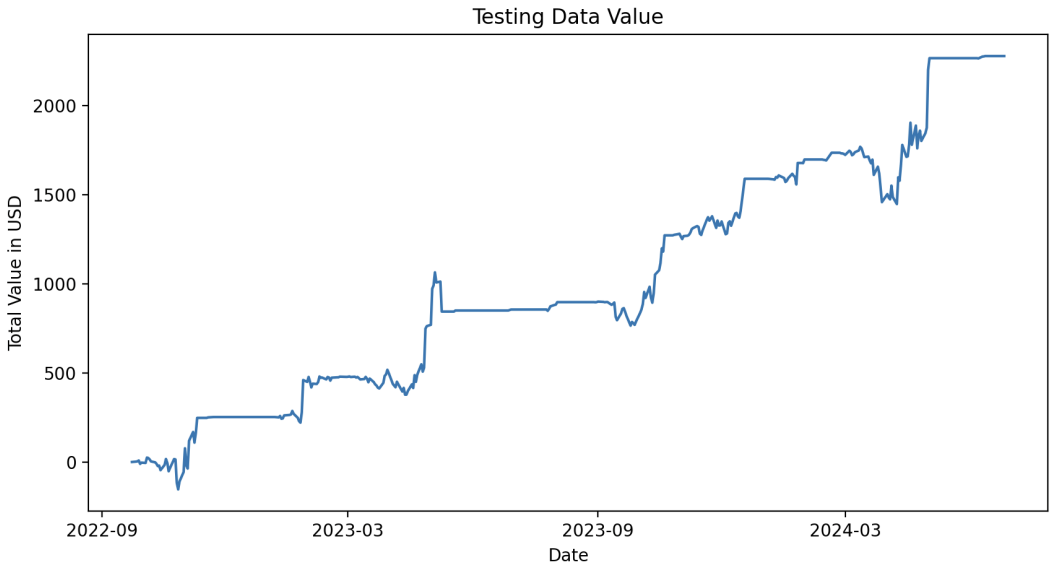
The training data generated a **profit of $5153.125** over a **period of approximately 3 years and 4 months**, as shown in the following images.



```
Training data profits: $5153.1249730719965
```

Similarly, the **testing data z-score distribution** and its **profits over time** is shown in the next 2 images. The testing data generated a **profit of $2277.206** over a **period of approximately 1 year 8 months**. This also shows that the model **works on new data.**






```
Testing data profits: $2277.205892932302
```

4 Limitations

1. it is assumed that the **cost of a transaction** in the stock market is **0**.
2. We can **buy/sell a fraction number of shares** to perfectly hedge the trade.
3. The amount of **data that is trained must be updated continuously** to avoid underfitting.
4. The strategy is **sensitive to disruptions that take place only in one share**.
5. Since we are creating a perfectly hedged trade, we need to **short and long simultaneously**.

5 Future Works

1. Take decisions on the **amount of risks** the model needs to take for a pair, **based on the confidence of preliminary tests**.
2. Continuously optimize market signals to generate more profit.