QF635 Market Microstructure and Algorithmic Trading

Group Members:
Song Yang
Liu Yue
Xia Tian
Gou Xinwen
Xia Qinru
Hao Xuanheng

1. Introduction

Pairs Trading is a strategy that seeks to identify short-term pricing anomalies between assets and capitalize on the convergence of their prices. It is essentially a statistical arbitrage play, betting that the spread between a pair of assets will likely revert to zero and involves taking a long position in the asset that is assessed to be undervalued vis-a-vis another asset, while simultaneously taking a short position in the overvalued. This strategy hinges on two key assumptions – first, that the assets in the pair are cointegrated and their spread exhibits mean-reversion characteristics; and second, that this relationship will continue to hold throughout the trading period.

In the classical Pairs Trading strategy, the search space for pairs is most typically restricted to equities. It is also often the case that the two stocks are chosen from the same sector e.g., PepsiCo Inc. (PEP) vs. Coca-Cola (KO) or Exxon Mobil (XOM) vs. Chevron (CVX) as these securities tend to be driven by similar economic fundamentals and thus are more likely to have prices that move in step with each other. Taking mutually offsetting positions in the pair aims to isolate the arbitrage opportunity while remaining sector-neutral, subject to the appropriate hedge ratios being implemented. However, as the intra-sector Pairs Trading strategy for equities becomes more widely implemented, it might be more susceptible to alpha decay, consequently resulting in a decrease in potential profits that can be gained. Besides the selection of suitable pairs, the entry and exit points of the trade are usually set at fixed thresholds, expressed as a factor of the spread's deviation from its historical average. More complex and targeted strategies may incorporate fine-tuning of these thresholds for each pair or implement rolling windows to adjust the entry and exit points on the fly.

In this project, our aim is to leverage the pairs trading strategy to identify and exploit pricing inefficiencies in cryptocurrency markets. The objectives of the project are twofold: First, to identify cointegrated pairs by conducting a thorough analysis to identify pairs of cryptocurrencies that exhibit strong cointegration. This involves statistical testing to ensure the pairs have a stable long-term relationship despite short-term deviations. Second, to implement and compare our trading strategy and stop-loss trading strategy to evaluate their effectiveness.

For the whole period cointegration analysis, we will analyze the cointegration over a comprehensive period (2020-2023) to identify the best pairs for trading. Two stop-loss strategies will be tested within this framework: a continuous return drop stop-loss, a stop-loss mechanism that triggers if the return drops continuously over a specified period, such as 4 hours, and a Bollinger Band stop-loss, a dynamic stop-loss strategy based on Bollinger Bands, which adjusts according to market volatility. For the yearly cointegration analysis, we will perform cointegration analysis on a two-year basis to identify pairs with potentially stronger relationships in shorter time frames. Similar to the whole period analysis, two stop-loss strategies will be tested: continuous return drop stop-loss and Bollinger Band stop-loss.

By comparing these strategies, we aim to determine which approach provides the most robust and profitable trading signals. This project will use historical data for various cryptocurrency pairs, sourced from reputable financial data providers, and employ statistical tools and trading algorithms for backtesting and performance evaluation. The primary tools and libraries used include Python, Pandas, Numpy, Statsmodels, and Matplotlib. Through this project, we seek to validate the effectiveness of pairs trading strategies in the cryptocurrency market and explore the impact of different stop-loss mechanisms on trading performance, ultimately aiming to provide insights into the optimal design of market-neutral trading strategies.

2. Data Processing

2.1 Data Acquisition

To conduct our pairs trading analysis, we acquired historical price data from Binance, focusing on various cryptocurrency pairs. This data includes hourly closing prices over a period from January 2020 to May 2024. Using a secure method to access the Binance API, we fetched the necessary historical data to form the foundation of our analysis.

2.2 Data Cleaning

After acquiring the raw data, the next step was data cleaning. This involved several key tasks:

1. Timestamp Conversion: The raw data included timestamps in milliseconds, which were converted to a readable datetime format. This conversion allowed us to work with date indices effectively.
2. Setting Index: We set the datetime as the index of our DataFrame to facilitate time series analysis.
3. Handling Missing Values: We checked for any missing values in the dataset. Any missing data points were appropriately handled either through interpolation or forward-filling techniques to ensure continuity in the time series.

4. Data Filtering: We filtered the data to ensure that only relevant columns, such as closing prices, were retained for the analysis. This step streamlined the dataset, focusing on the key metrics required for the pairs trading strategy.

2.3 Data Processing

Once the data was cleaned, we proceeded with data processing to prepare it for analysis:

1. Normalization: To compare different cryptocurrency pairs effectively, we normalized the closing prices. This step ensured that price movements were on a comparable scale, facilitating accurate statistical analysis.
2. Visual Inspection: We plotted the closing prices over time to visually inspect the data for any anomalies or trends. This initial visualization helped in understanding the overall price movement and identifying any irregular patterns that might require further investigation.
3. Cointegration Testing: Using statistical tools, we tested for cointegration between different pairs of cryptocurrencies. Cointegration testing was a crucial step to identify pairs that exhibited a stable long-term relationship despite short-term price deviations.

2.4 Split Data

The dataset was divided into four main subsets representing different periods: one covering from January 1, 2020, to December 31, 2021; the second from January 1, 2021, to December 31, 2022; the third from January 1, 2022, to December 31, 2023; and the fourth from January 1, 2024 to May 31,2024. The first three are training sets to test cointegration and choose the optimal pairs, and the last one is the test set.

For more detailed analysis, the training data was further divided into single-year subsets: one covering from January 1, 2021, to December 31, 2021; another from January 1, 2022, to December 31, 2022; and a third from January 1, 2023, to December 31, 2023. Then, we calculate OLS regression for each single-year subset. For the accuracy of variables, we absorb new information and timely discard the old, and select the rolling window of 8760 hours (365 days) to compute the rolling intercept, rolling beta and rolling spread. For example, we use the entire year of 2020's data to perform an OLS regression, and the resulting beta and intercept are used as indicators to calculate the spread and position at 1 AM on January 1, 2021. Then, we roll the data to calculate each subsequent indicator.

This comprehensive splitting allowed for robust backtesting and validation of the trading strategies under different market conditions, ensuring the strategies' effectiveness and resilience across various time periods.

3.Research Results

3.1.Test Cointegration

Cointegration is a statistical property of a collection of time series variables which, when combined in a certain way, exhibit a stable long-term relationship. In the context of pairs trading, cointegration is used to identify pairs of assets whose prices tend to move together over time. This is particularly useful because if two assets are cointegrated, the spread between their prices should be stationary, meaning it will revert to a mean over time. This reversion property forms the basis of pairs trading strategies, where traders can exploit deviations from the mean spread to generate profits.

We divided the analysis into three separate periods and performed cointegration tests respectively. This allowed us to evaluate cointegration across different market conditions.

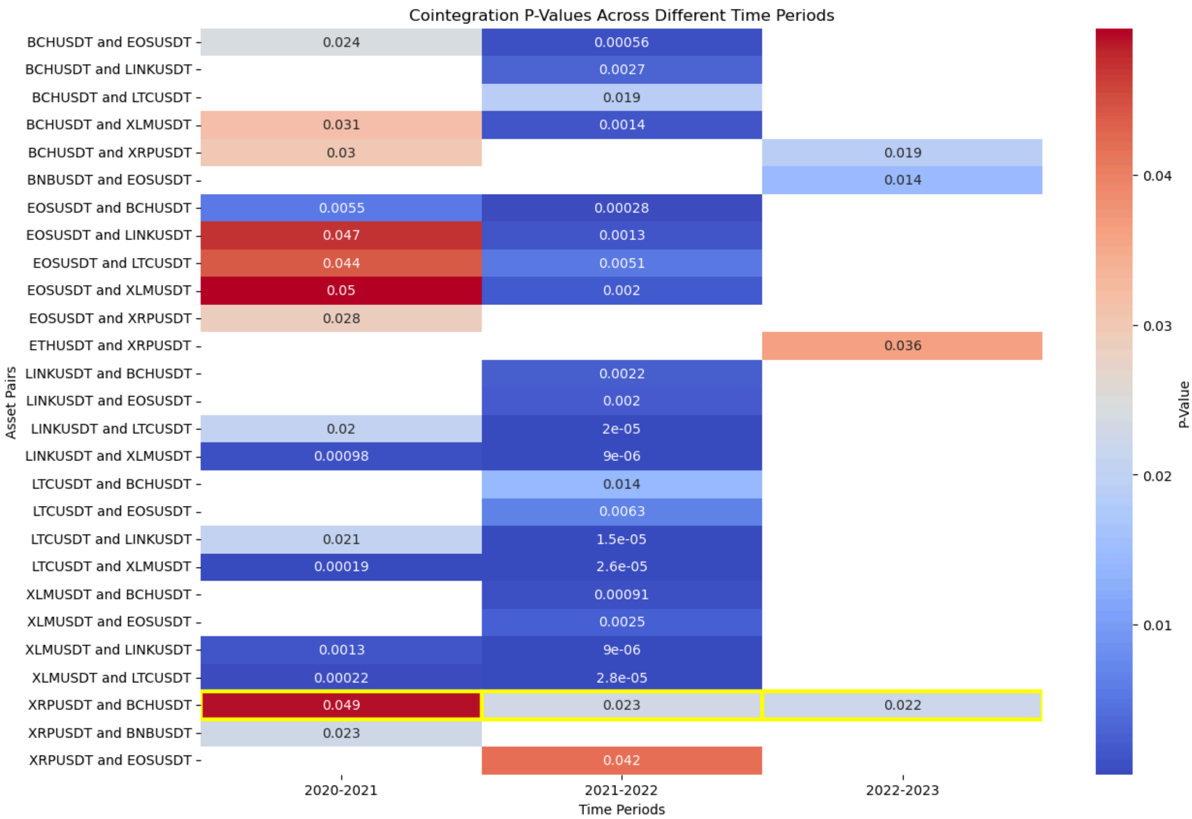


Figure 3.1 Cointegration P-Value Heatmap

The cointegration test results are summarized in the heatmap, which shows the p-values for different asset pairs across three time periods. A p-value threshold of 0.05 was used to determine significant cointegration relationships. Lower p-values suggest stronger evidence of cointegration.

Based on the cointegration tests, we identified that the pair XRPUSDT and BCHUSDT consistently exhibited significant cointegration across all three time periods. This indicates a stable long-term relationship, making it a suitable candidate for our pairs trading strategy.

3.2.Trading Strategy

Choosing the only pair that passes the cointegration test, XRPUSDT and BCHUSDT, we begin to design our strategy. Before proceeding, we have already split the hourly data into four datasets, including three train datasets and one test dataset. For each dataset, we apply our trading strategy and observe the relevant metrics.

The concrete strategy is as follows:

Firstly, we perform OLS regression to obtain the values of beta and intercept, then use these values to calculate the spread.

$$XRPUSDT = intercept + beta * BCHUSDT + \varepsilon$$
$$Spread = XRPUSDT - intercept - beta * BCHUSDT$$

To make sure the accuracy of variables, absorb new information and timely discard the old, we select the rolling window of 8760 hours (365 days) to compute the rolling intercept, rolling beta and rolling spread.

Secondly, we use the rolling window of 90 hours to calculate the rolling mean of spread and standard deviation of spread, which are then used to normalize the spread.

Thirdly, we set the entry condition at 1.96 ,the exit condition at 0.25 and the minimum threshold value at 0.005 to determine our signals. The first two conditions pertain to the normalized spread, while the last condition relates to the absolute spread. When the normalized spread exceeds 1.96 and the absolute spread is more than 0.005 , we short the spread, anticipating that the high value will revert to zero. Conversely, when the normalized spread is less than -1.96 and the absolute spread is more than 0.005, we long the spread. When the absolute value of the normalized spread is less than 0.25, we close the position.

After setting the signals, we use forward fill to complete the XRPUSDT positions. For the BCHUSDT positions, we use the rolling beta to calculate them.

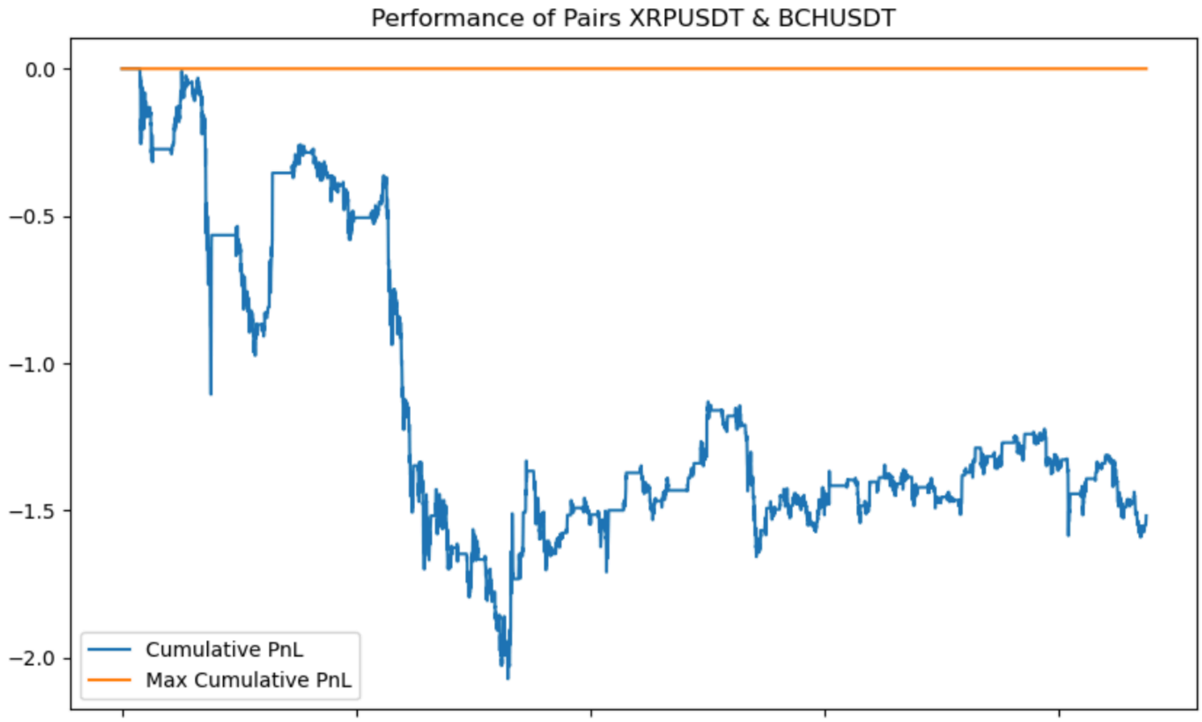
$$BCH\,position = XRP\,position * beta * (-1)$$

Finally, we compute the log returns of two cryptocurrencies to obtain the hourly PNL and cumulative PNL of our strategy.

$$Hourly\,PNL = XRP\,position * XRP\,return + BCH\,position * BCH\,return$$

3.3. Backtesting result

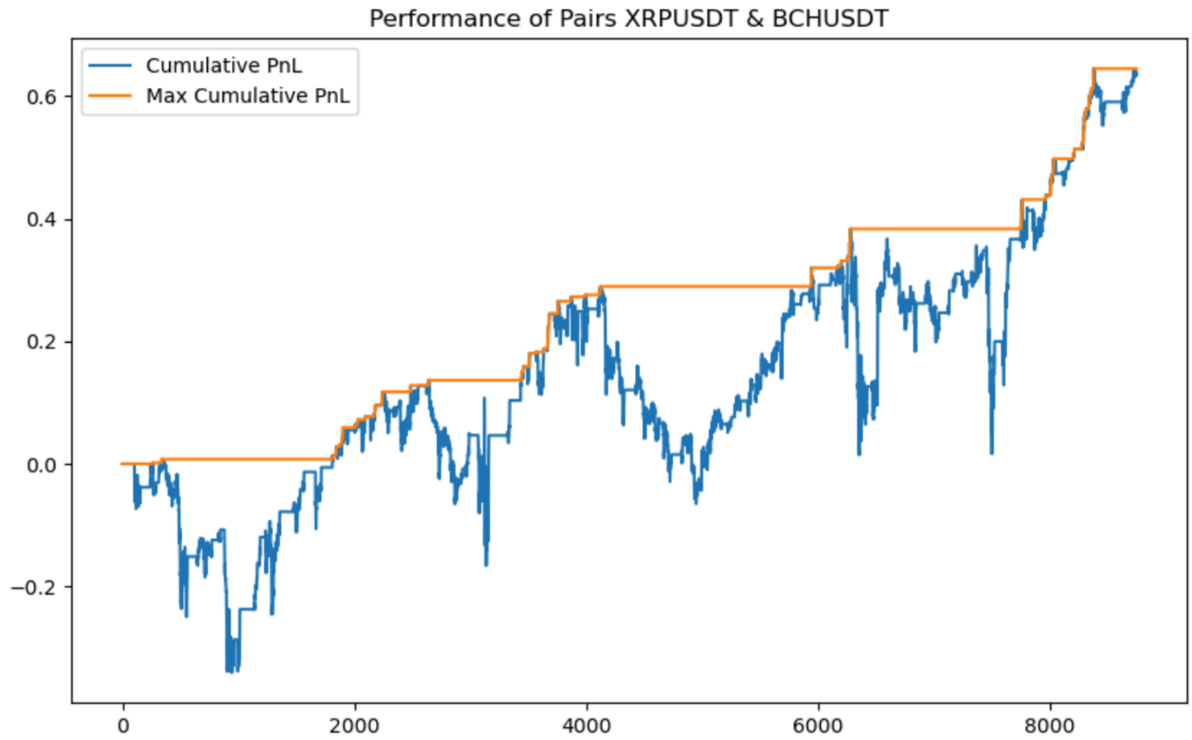
For the strategy performance on the first train dataset (2021-01-01 to 2021-12-31), the graph of cumulative return and relevant metrics are as follows:



	Year	Stop-Loss	Sharpe Ratio	Average Annual Returns (%)	Total Returns (%)	Max Drawdown (%)	Max Drawdown Period
Train	2021	No	-1.017756	-104.894237	-151.80874	-207.200261	-138 days +07:00:00

Figure 3.3a Strategy Performance in Year 2021

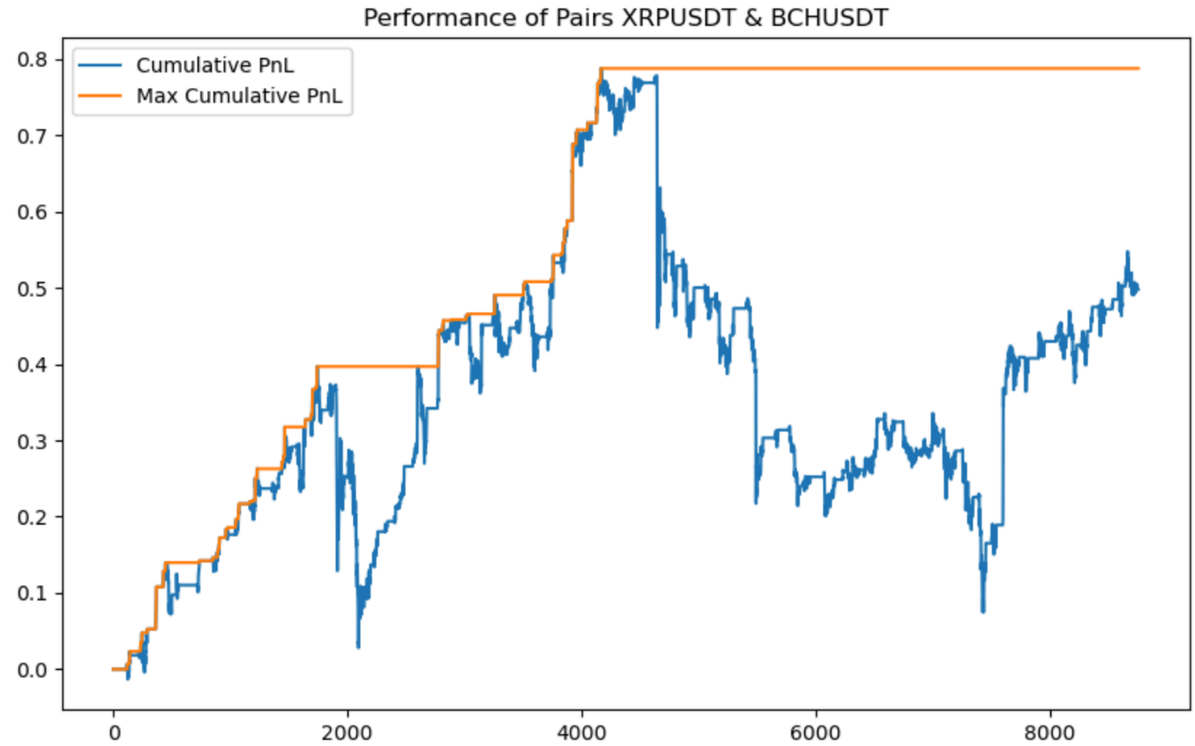
For the strategy performance on the second train dataset (2022-01-01 to 2022-12-31), the graph of cumulative return and relevant metrics are as follows:



	Year	Stop-Loss	Sharpe Ratio	Average Annual Returns (%)	Total Returns (%)	Max Drawdown (%)	Max Drawdown Period
Train	2022	No	0.756918	44.046209	63.702412	-36.886213	-265 days +07:00:00

Figure 3.3b Strategy Performance in Year 2022

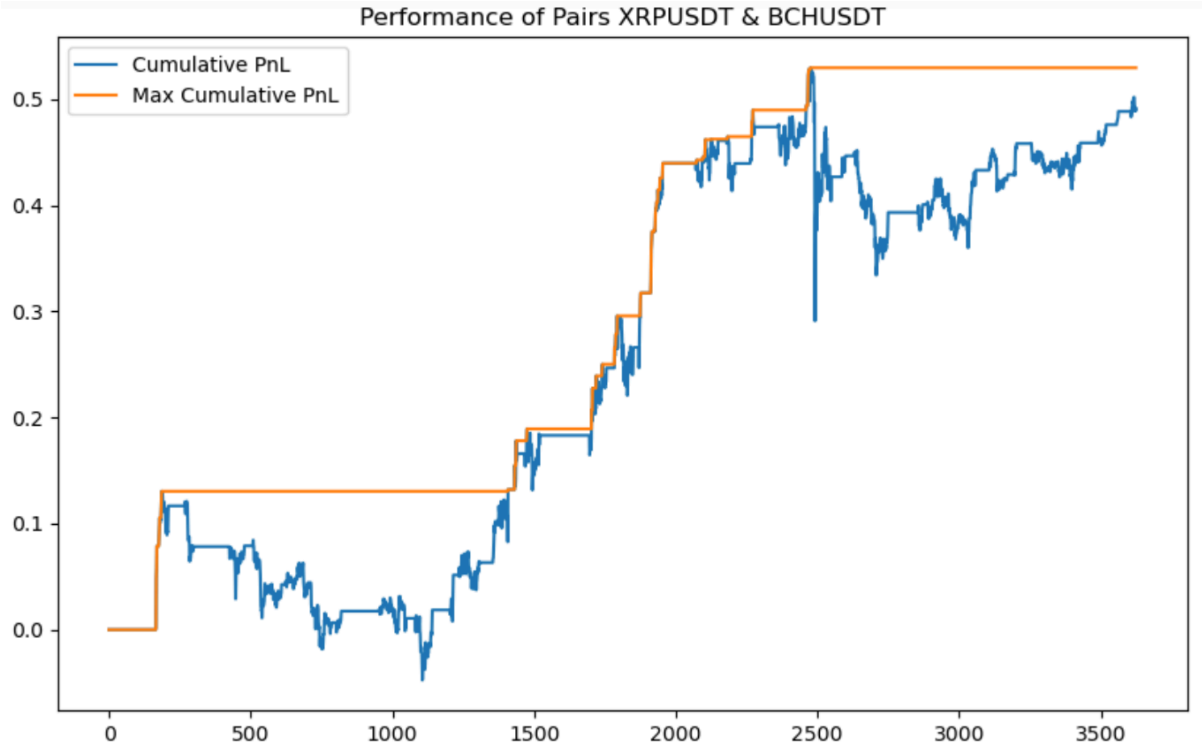
For the strategy performance on the third train dataset (2023-01-01 to 2023-12-31), the graph of cumulative return and relevant metrics are as follows:



	Year	Stop-Loss	Sharpe Ratio	Average Annual Returns (%)	Total Returns (%)	Max Drawdown (%)	Max Drawdown Period
Train	2023	No	0.705125	34.370759	49.777361	-71.267322	-310 days +03:00:00

Figure 3.3c Strategy Performance in Year 2023

For the strategy performance on the test dataset(2024-01-01 to 2024-05-31), the graph of cumulative return and relevant metrics are as follows:



	Year	Stop-Loss	Sharpe Ratio	Average Annual Returns (%)	Total Returns (%)	Max Drawdown (%)	Max Drawdown Period
Test	2024	No	1.931665	81.955289	49.108129	-23.850902	-104 days +05:00:00

Figure 3.3d Strategy Performance in Year 2024

4. Risk Control

4.1.Execution Risk

The backtest results are based on the hourly closing prices, representing the matched prices between buyers and sellers. In a live trading scenario, however, we must decide between using limit orders or market orders. Since pairs trading requires simultaneous execution of two assets, market orders seem like a straightforward choice. Nevertheless, if prices fluctuate significantly, market orders could lead to substantial losses. Conversely, with limit orders, there's a risk of only one side of the trade being executed. This limitation highlights a challenge in our strategy that requires further investigation.

Additionally, due to the low OLS beta between XRPUSDT and BCHUSDT, we have scaled the quantity by 1000 times to ensure the BCHUSDT order size exceeds the minimum threshold. However, this approach also increases the quantity of XRPUSDT by 1000 times, potentially leading to liquidity risks and execution issues in the market.

4.2.Model Risk

The pairs trading strategy primarily relies on two key factors: cointegration and spread. First, we need to assess the cointegration between our two assets, acknowledging that this relationship may vary over time. Second, we need to construct a dynamic spread that accurately reflects temporal changes. Third, given the time-varying nature of the spread, the Z-score signal must also be adaptable. To address these considerations, we test for cointegration using a two-year rolling window to ensure stability within the training set. We use rolling beta in OLS regression to determine the spread and calculate the Z-score based on the rolling mean and standard deviation of the spread.

Below are the results when risks are not managed as previously outlined: Using data from 2020 to 2023 to test for cointegration, we found that LINKUSDT and XLMUSDT had the highest correlation, with a P-value of 0.000097. However, the P-value for these assets in 2023 was 0.381421, indicating a lack of significant cointegration and suggesting they might not be suitable for pairing. Therefore, our strategy employs a two-year data window to test for cointegration, avoiding pairs that are not cointegrated in recent years. Figure 4.1 illustrates the price trends of LINKUSDT and XLMUSDT in 2023.

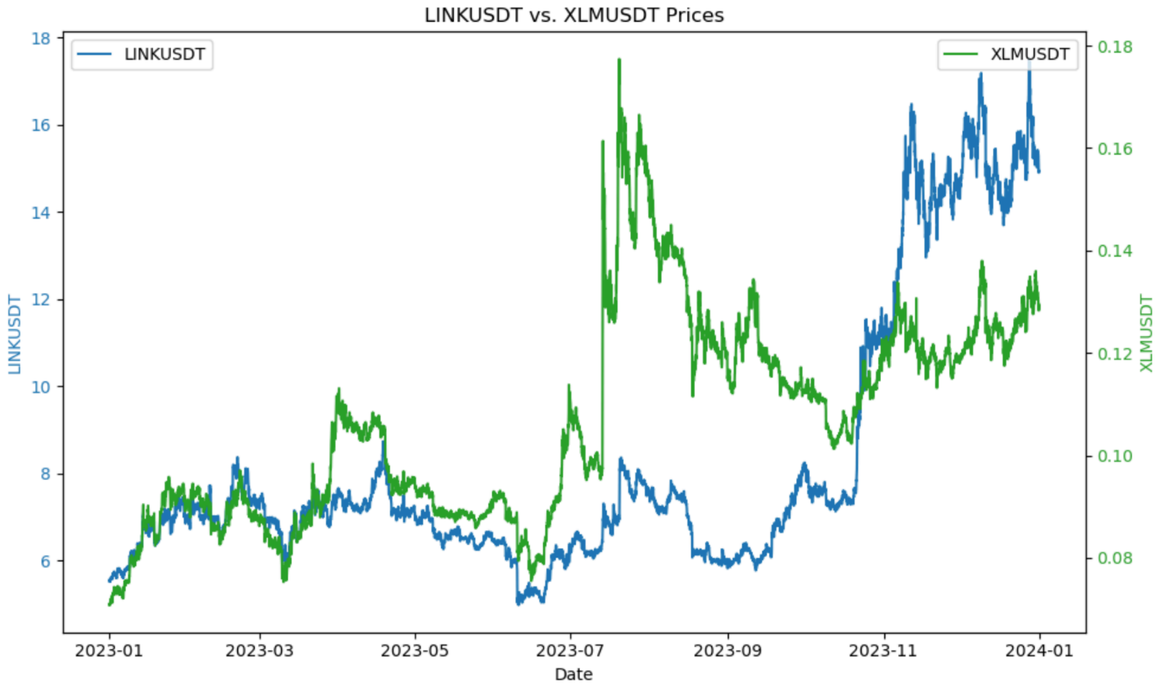


Figure 4.1 Prices of LINKUSDT and XLMUSDT

Besides, if we use the beta derived from 2022 and 2023 data to construct the 2024 spread, we will unsurprisingly find that the spread does not fluctuate around zero. In this case, our Z-score signal can not work well. Therefore, we use rolling beta to decide spread, which can give us a more reasonable result. Figure 4.2a shows the spread derived from non-rolling beta. Figure 4.2b shows the spread derived from rolling beta.

Figure 4.2a Spread of XRPUSDT and BCHUSDT using non-rolling beta

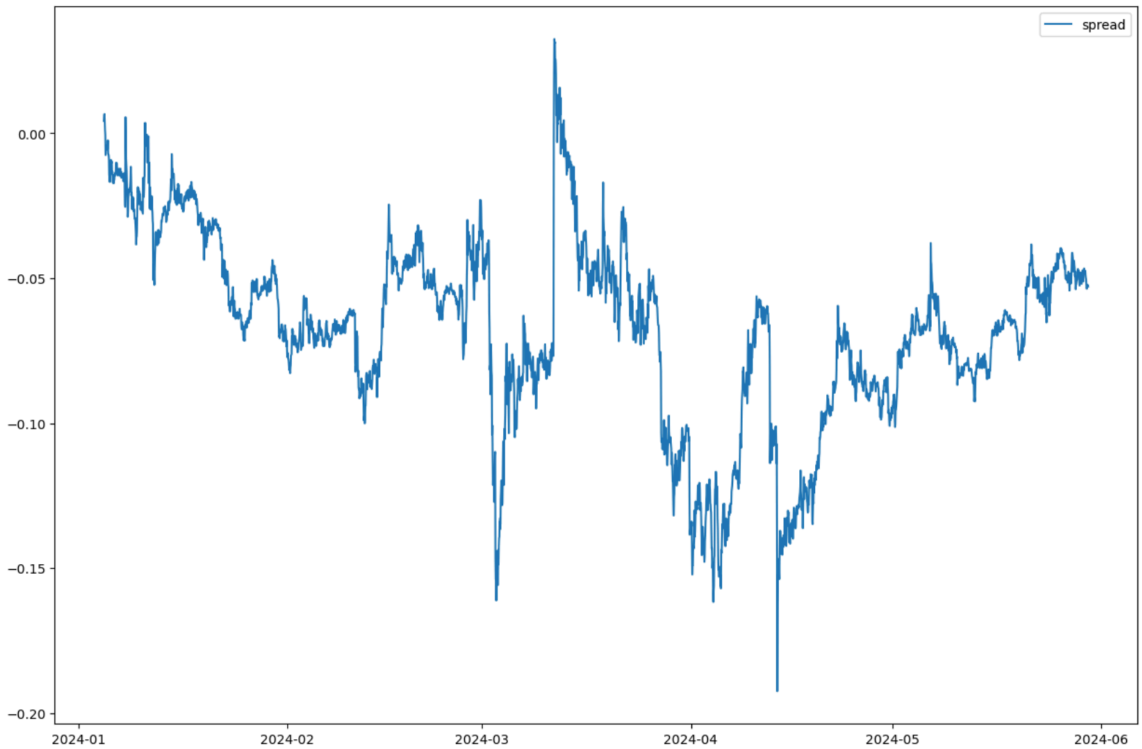


Figure 4.2b Spread of XRPUSDT and BCHUSDT using rolling beta

4.3. Market Risk

The most important assumption of our strategy is that the spread is mean-reversion. However, the spread might deviate from the mean a lot of range, which can cause loss. Therefore, we need to add a stop-loss strategy. When the unrealized loss in our balance is over the 1.96 standard deviation of the rolling PNL, we adjust the signal to zero to stop loss. Figure 4.3a shows the signal without stop-loss. Figure 4.3b shows the signal with stop-loss.

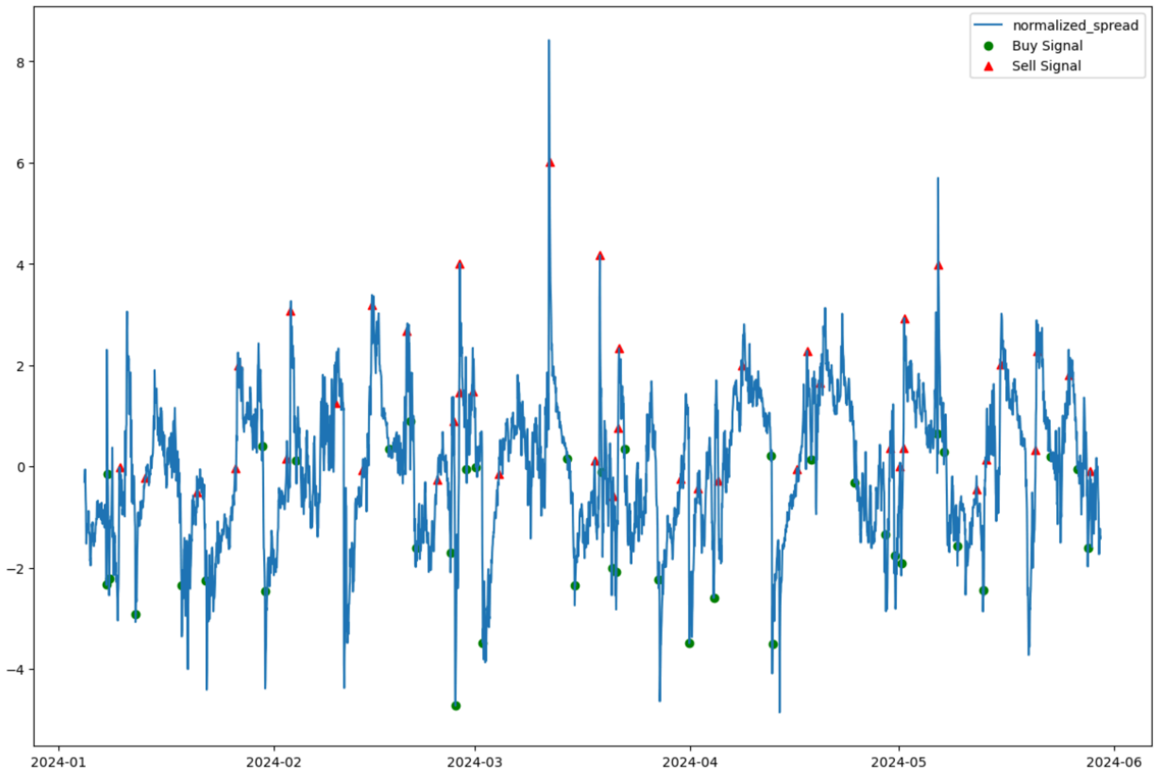


Figure 4.3a Signal of pairs trading(XRPUSDT and BCHUSD) without stop-loss

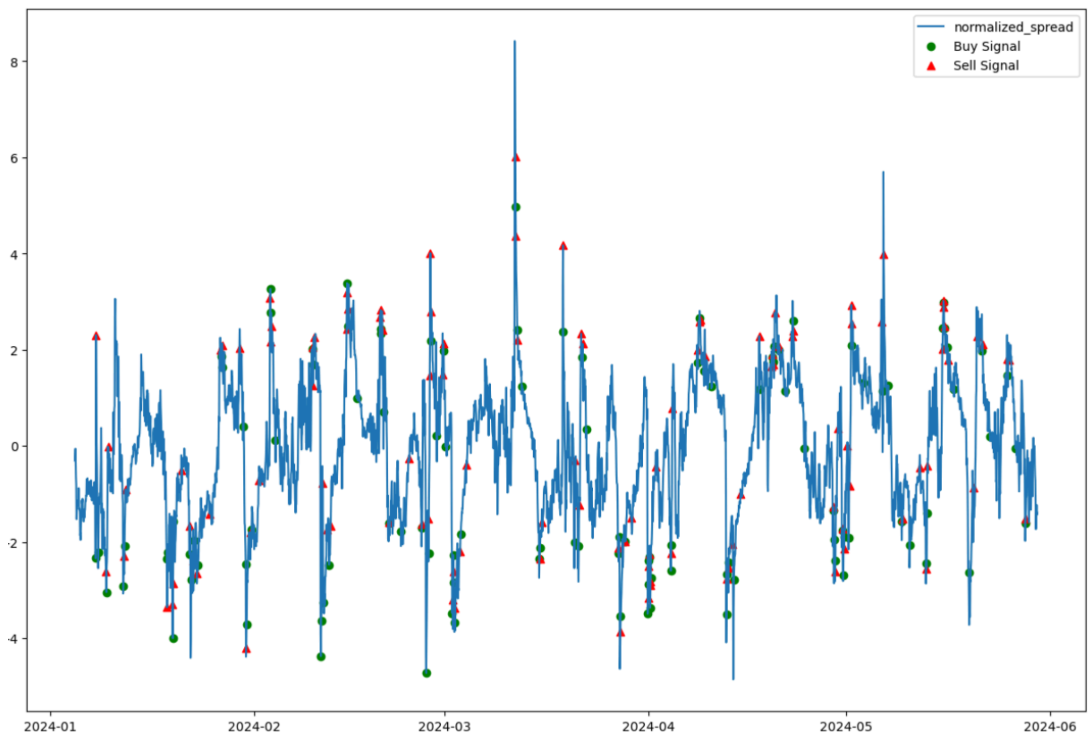


Figure 4.3b Signal of pairs trading(XRPUSDT and BCHUSD) with stop-loss

After using stop-loss strategy, the annualized sharpe ratio and annual return increase significantly in train data. Meanwhile, the max drawdown and max drawdown period decreases significantly. Figure 4.3c shows the performance within the train dataset. Figure 4.3d shows performance without stop-loss in test data. Figure 4.3e shows performance with stop-loss in test data.

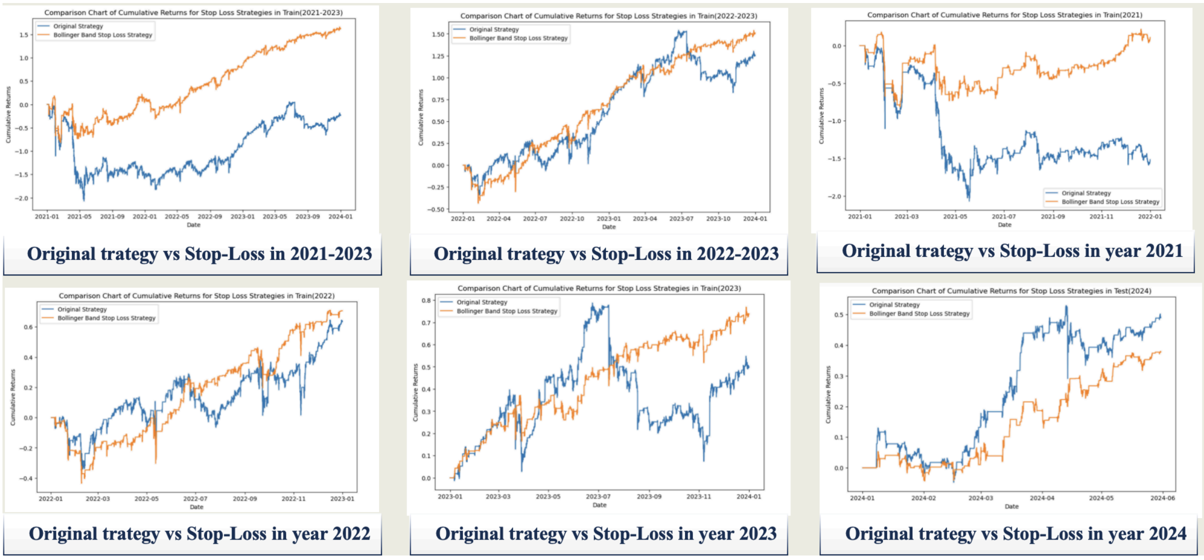
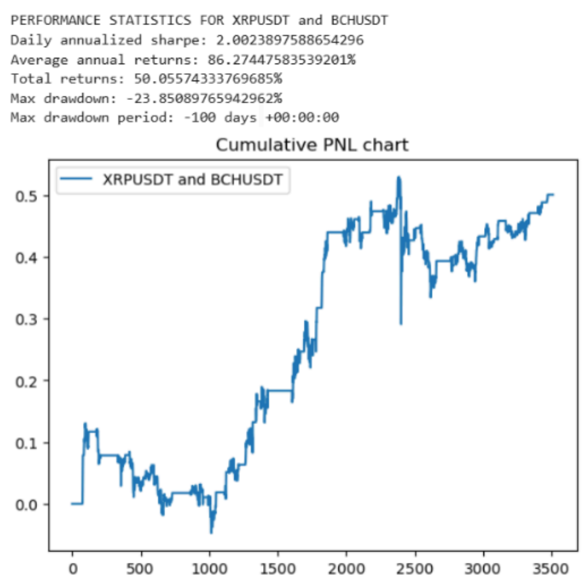


Figure 4.3c Performance of pairs trading (XRPUSDT and BCHUSD) within train data



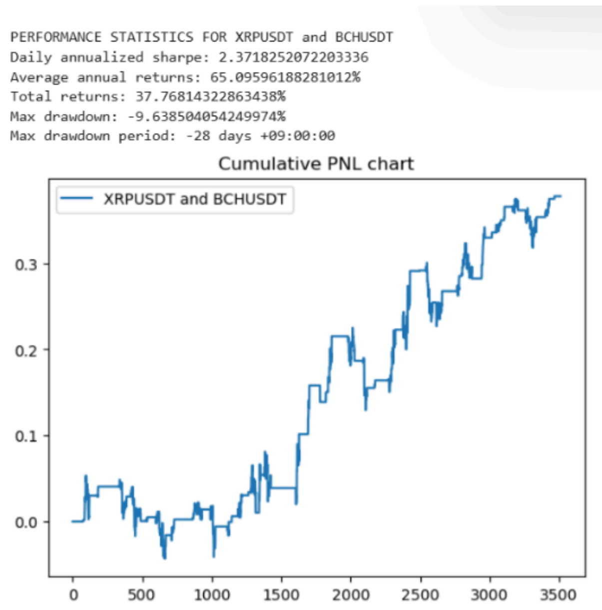


Figure 4.3d Performance of pairs trading (XRPUSDT and BCHUSD) without stop-loss

Figure 4.3e Performance of pairs trading (XRPUSDT and BCHUSD) with stop-loss

5. Conclusion

The table below is the results of our strategy with sharpe ratio, average annual returns, total returns, max drawdown and max drawdown periods.

	Year	Stop-Loss	Sharpe Ratio	Average Annual Returns (%)	Total Returns (%)	Max Drawdown (%)	Max Drawdown Period
Train	2021-2023	No	-0.072397	-5.352018	-23.248711	-207.200261	-138 days +07:00:00
Train	2021-2023	Strategy 1	-0.280951	-17.763218	-77.161916	-219.061716	-408 days +20:00:00
Train	2021-2023	Strategy 2	0.684103	37.506237	162.923920	-99.834676	-51 days +12:00:00
Train	2022-2023	No	0.804799	43.234076	125.141491	-71.267322	-675 days +16:00:00
Train	2022-2023	Strategy 1	0.682684	32.008314	92.648404	-54.134008	-679 days +23:00:00
Train	2022-2023	Strategy 2	1.352549	52.135847	150.907761	-43.863347	-38 days +08:00:00
Train	2021	No	-1.017756	-104.894237	-151.808740	-207.200261	-138 days +07:00:00
Train	2021	Strategy1	-1.373676	-119.694748	-173.228858	-216.807407	-173 days +05:00:00
Train	2021	Strategy2	0.096735	7.523169	10.887946	-99.834676	-51 days +12:00:00
Train	2022	No	0.756918	44.046209	63.702412	-36.886213	-265 days +07:00:00
Train	2022	Strategy1	0.280700	14.310007	20.696036	-51.304780	-42 days +15:00:00
Train	2022	Strategy2	1.100226	48.847690	70.646618	-43.863347	-38 days +08:00:00
Train	2023	No	0.705125	34.370759	49.777361	-71.267322	-310 days +03:00:00
Train	2023	Strategy1	0.984536	41.630078	60.290650	-54.134008	-314 days +10:00:00
Train	2023	Strategy2	1.614497	50.956881	73.798168	-20.683336	-88 days +15:00:00
Test	2024	No	1.931665	81.955289	49.108129	-23.850902	-104 days +05:00:00
Test	2024	Strategy1	1.928425	73.223252	43.875837	-16.383913	-104 days +05:00:00
Test	2024	Strategy2	2.341190	63.280218	37.917909	-9.638509	-32 days +14:00:00