QFin Pairs Trading Algorithm Report

Kane Alexander, Jonathan Choong, Jessie Xie, Kevin Shah Mansouri
June 16, 2021

---

1 Introduction

Pairs trading is a market-neutral trading strategy, which is based on the co-movement of two stock prices with respect to each other. If the prices of two stock diverge, taking a long position in one and a simultaneous short position in the other can be used to profit from the expected future re-convergence of the prices.

2 Methodology

There are three main steps in our strategy. First, potential stock pairs are chosen using a variety of qualitative and quantitative methods. Second, design and implementation of a pairs trading algorithm to work within the project constraints. Finally, backtesting the selected pairs on the trading algorithm to select the most promising pair of stocks. Parameters of the algorithm were restricted to S&P500 stocks with a 0.01% fee per trade and $20,000 starting capital.

2.1 Choosing Pairs

To reduce the net number of pair calculations needed to be undertaken the S&P500 constituents were divided into sectors. Stocks in the same sector tend to have stronger correlations than those in unrelated industries. Selecting pairs within the same sector reduces the likely-hood whereby simultaneous movement of stock prices is not supported by practical reasons, but simply due to mathematical randomness. Eight distinct sectors based on GICS sector classification were chosen, and all pairs within each sector were examined. The eight sectors include Communication Services, Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Materials, and Real Estate.

To find suitable pairs within these eight sectors, a preliminary correlation test for each potential pair was performed. Correlation is a statistic that measures the degree to which two variables move in relation to each other. The correlation coefficient is a number ranging from -1 to 1. A value of -1 means that two variables have a perfect negative correlation, a value of 0 means there is no statistical correlation, and a value of 1 means a perfect positive correlation. Pairs trading is an adaptive strategy that is viable on both highly positive and highly negatively correlated assets. The magnitude of each pairs correlation coefficient determined the top five most viable pairs within each sector.

Correlation describes the tendency for stocks to move in similar or opposite directions, but alone cannot determine if a pair is suitable for pairs trading. A suitable pair must not only move in the same direction but also consistently converge to the same relative price. This was tested by applying the augmented Engle-Granger two-step cointegration test on each of the viable pairs. Cointegration test identifies scenarios where two non-stationary time series are integrated together in a way that they cannot deviate from equilibrium in the long term. The Engle-Granger Two-Step method starts by creating residuals based on the static regression and then testing the residuals for the presence of unit roots. It uses the Augmented Dickey-Fuller Test (ADF) to test for stationary units in time series. If the time series is cointegrated, the Engle-Granger method will show the stationarity of the residuals. The null hypothesis is that there is no cointegration. If the P-values are smaller than 0.05, we will reject the null hypothesis at the 5% significant level and keep these pairs for backtesting.

2.2 Algorithm

The algorithm implements a cointegration method adapted from QuantConnect [1] to accommodate for stock data of minute resolution. Using a training time and rolling window size of 1 month, linear regression was performed on the prices of the two stocks (x, y) to determine spread. This was done using equation 1, where α is the regression intercept and β is the cointegration coefficient.

$$spread = \log(price_y) - \beta \cdot \log(price_x) - \alpha \tag{1}$$

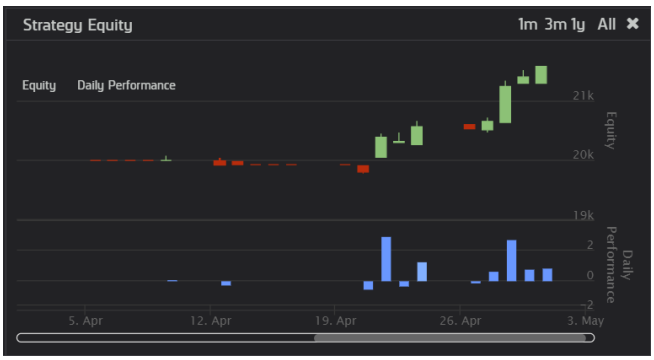
Using the spread for the current rolling window, the mean and standard deviation could be calculated which would provide reference values for triggering trading signals. Assuming the entire portfolio has yet to be invested, the algorithm would allocate 50% of the portfolio to buying each of the two stocks. Since a threshold of 2 standard deviations

was used, if the current spread decreased to over -2 standard deviations from the rolling window mean, the algorithm would take a long position on stock x and short on y. However, if the current spread was more than 2 standard deviations from the mean, it would take a short position on x and long on y.

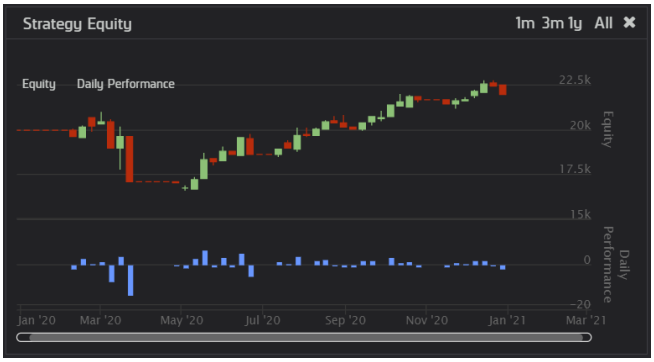
If the portfolio has been entirely invested, the algorithm will either hold or liquidate. If the holdings of stock x is short and the spread has dipped below mean, or if it is long and the spread has increased above the mean, this is an indicator that the two stocks have mean reverted, which triggers a liquidation of the portfolio and theoretically should return a profit.

3 Results

The initial correlation testing within industries resulted in a total of 40 viable pairs, which reduced to 21 after conducting cointergration tests. Each pair was subsequently tested over the 01/04/2021 to 01/05/2021 period with the pair with the largest P/L ratio being selected for submission. "NOV" and "HAL" were chosen as the most profitable pair with a total profit of 7.86% over the month. This test was increased over the 1/1/2020 to 1/1/2021 to test for longer term reliability and returned a net profit of 9.83% over the year. These stocks from the energy sector didn't have the strongest correlation coefficient (0.9896) but had reliable cointergration (0.009747) that lead to optimal pairs trading.



(a) NOV and HAL pair results (1/4/2021 - 1/5/2021)



(b) NOV and HAL pair results (1/1/2020 - 1/1/2021)

Figure 1: Using QuantConnect to backtest the implemented pairs trading algorithm for stocks NOV and HAL over the year of 2020.

4 Conclusion

The pairs trading algorithm has proven to be a market neutral strategy which serves to make profit from both increasing and decreasing prices in the market. It is a strategy that minimises significant losses but can also restrict the size of profits. Potential improvements for the algorithm could include dynamic order sizing to reduce the toll of fees, where portfolio allocation could be adjusted using the cointegration coefficient or other methods. Another improvement could be the implementation of Kalman filters to dynamically update the hedge ratio and reduce the randomness of trading signals. A copula method could also be used in place of the cointegration method, which introduces copula functions to select correlated stocks and generate trading signals. The copula method has been mentioned by many articles to give a better return than the cointegration method.

References

[1] QUANTCONNECT. Pairs Trading-Copula vs Cointegration [Online]. Available at: https://www.quantconnect.com/tutorials/strategy-library [Accessed 10/06/21].

[2] QUANTCONNECT. Tutorials - Strategy Library - Pairs Trading with Stocks [online] Available at: <https://www.quantconnect.com/tutorials/strategy-library/pairs-trading-with-stocks> [Accessed 10 June 2021].

[3] MEDIUM 2021. Pairs Trading using Data-Driven Techniques: Simple Trading Strategies Part 3. [online] Available at: <https://medium.com/auquan/pairs-trading-data-science-7dbedafcfe5a> [Accessed 1 June 2021].

[4] STATSMODELS 2021. statsmodels.tsa.stattools.coint — statsmodels. [online] Available at: <https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.coint.html?fbclid=IwAR3AcAajxo6MIGejzBsmyhry8a3OySGzRhvPXqaiaYGjNhEJJ_IDpzKUSHE> [Accessed 5 May 2021].