Pairs Trading Analysis of the S&P 500

Gabby Rogers

Department of Computer Science, Trinity College, Hartford, CT

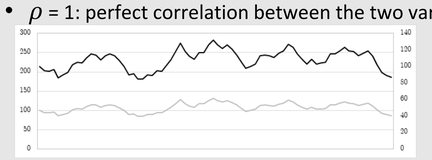
Introduction

Pairs Trading is a low-risk trading strategy in which you find two assets that share an underlying economic link and have a historically constant price spread. From time to time, this price spread temporarily diverges due to temporary supply/demand shifts, large buy/sell orders for one security, reaction for important news about one of the companies, etc. In this scenario, one stock moves up while the other moves down relative to each other. You can make a pairs trade if you expect this divergence to revert back to normal with time. The pairs trade would be to sell the outperforming stock and to buy the underperforming stock.

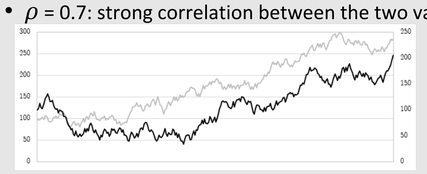
Correlation

Correlation (ρ) measures the degree of the linear relationship between variables:

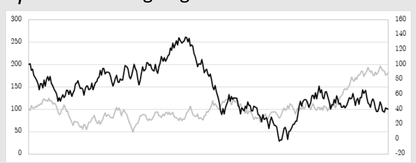
- ρ = 1: perfect correlation between the two variables



- ρ = 0.7: strong correlation between the two variables



- ρ = -0.7: strong negative correlation between the two variables

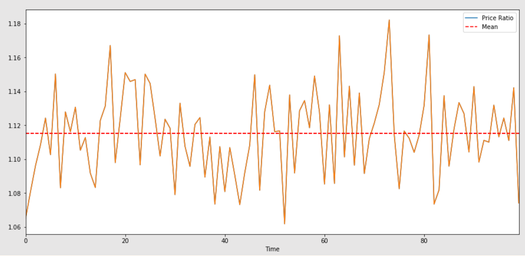


Correlation helps identify short-term relationships among variables. However, it is not used to explain long-term relationships among financial data.

Cointegration

Cointegration means that the ratio between two series will vary around a mean. For pairs trading to work between two timeseries, the expected value of the ratio over time must converge to the mean, i.e. they should be cointegrated.

The graph below is the plot of the ratio between prices of two cointegration stocks and its mean:



Data and Methods

Data

- The data used for this project is constituents of the S&P500 Index gathered from Yahoo Finance
- The timespan of the data ranges from 1-1-2011 to the current date.
- Note: Given the effect of the COVID-19 pandemic on the stock market, the time period from mid-February 2019 until June 2020 was omitted. During this time the market was experiencing sharp uptrends and downtrends that threw asset pair relationships into an abnormal tailspin.
- From here, lists of stocks were selected based on economic links such as market capitalization, sector, and product.

Methods:

For each selected list of stocks, I performed the Engle-Granger two-step cointegration test.

- Null Hypothesis: No cointegration exists
- Alternate Hypothesis: Cointegration exists
- Returns the t-statistic of unit-root test on residuals.
- MacKinnon's approximate, asymptotic p-value.
- Critical values for the test statistic at the 1%, 5%, and 10% levels based on regression curve.

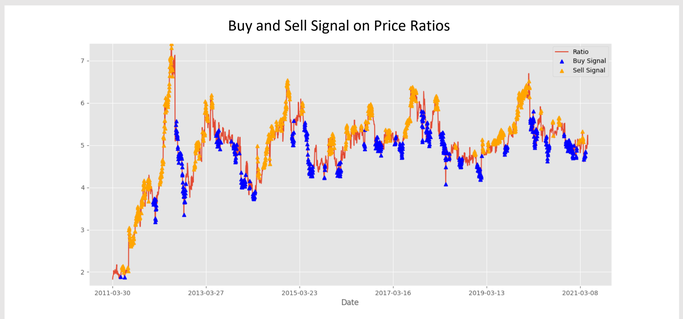
Creating the Trading Strategy

Split Data:	Feature Engineering:	Model Selection	Train: Plot ratios & buy/sell signals from the z-score Validate: Plot prices & buy/sell signals from the z-score Optimize: adjust windows, change thresholds, etc.	Back-Test: Calculate Profit & Loss on test data
• Training Range: 7 years • Testing Range: 3 years	• 60-day MA of Ratio • 5-day MA of Ratio • 60-day SD	• Buy (1) whenever z-score < -1 • Sell (-1) whenever z-score > 1		

Procedure:

Split Data → Feature Engineering → Model Selection → Train, Validate, and Optimize → Back-test

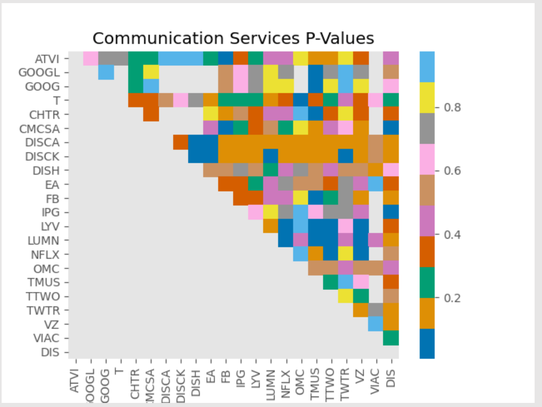
The graph below represents the Train, Validate, and Optimize step. This displays the buy and sell signals from the z-score:



We can see that we sell the ratio at the orange triangles and buy it back at the blue triangles.

Preliminary Results

The chart below is a heatmap displaying the p-values resulting from performing the Engle-Granger Test between each pair of stocks in a list of companies in the Communication Services Sector:



Pairs with a significant p-value are represented by a dark blue square.

In this list, pairs with a p-value < 0.02 include:

- CHTR and TMUS
- CMCSA and FB
- NFLX and TTWO

This process was also used on other small lists of data such as:

- Technology companies
- Companies with a large market capitalization

Avenues for Further Work

In practice, trading algorithms may implement more sophisticated statistical methods such as:

- Kalman Filter
- Half-life of mean reversion inferred from an Orstein—Uhlenbeck process
- Hurst exponent

This analysis could also be expanded into a real-time trading algorithm.

References

Abhi. (2020, September 22). Quant Trading : An Introduction To Pairs Trading. Retrieved from https://towardsdatascience.com/quant-trading-an-introduction-to-pairs-trading-5ce50c03177e

Auquan. (2020, March 05). Pairs Trading using Data-Driven Techniques: Simple Trading Strategies Part 3. Retrieved from https://medium.com/auquan/pairs-trading-data-science-7dbedafcfe5a

Acknowledgements

This project was supported by:

- Chris Armen, Trinity College
- Krishna Venugopal, Olivine Research