ORIE 5741 Class Project Proposal

Amanda Zou (yz2849), Linjia Feng (lf433), Jingde Wan (jw2379)

March 17, 2023

Topic Overview

Pairs trading is a classical statistical arbitrage investment strategy that was developed in the 1980s. The main idea of pairs trading is to identify two highly correlated assets and make trading signals when a statistical anomaly is detected in the spread between the two assets. This is based on the belief that the two assets' prices, which have moved closely together in the past, will continue to do so consistently in the future.

Questions to explore

The successful pairs trading strategy relies on two things: finding the right pairs and designing a robust trading model/mechanism to open/close market position. However, traditional methods are imperfect. On the one hand, the cost of trying each pair of assets in the universe is quite expensive using traditional methods. On the other hand, traditional pairs trading implementing a threshold-based trading model is lack of robustness. Fortunately, by using machine learning techniques, we are more likely to advance. We would like to explore these two questions of pairs trading:

- Can unsupervised learning like clustering help us to find promising clusters that contain profitable pairs in the universe of assets?
- Can supervised learning like regression combined with time series forecasting help us to generate a forecasting-based trading model on spread to achieve more robust performance?

Data

We would like to choose Exchange-Traded Funds (ETFs) as the asset class in this project. ETFs are a type of security that tracks a particular index, sector, commodity, or other asset and can be purchased or sold on a stock exchange in the same way that a common stock does. One advantage that ETF has is that ETF is more robust over time, which means that the past and future behaviors of the same ETF will be more consistent. We will collect data from Yahoo Finance API.