Pairs Trading Algorithm

Motivation

Our group decided to design a pairs trading investment strategy for this project. Pairs trading is a statistical arbitrage strategy that has been implemented by firms such as D.E. Shaw, Cornerstone, and Morgan Stanley. The first step of a pairs trading strategy is identifying a pair of stocks that are both correlated and cointegrated. The prices of a pair of correlated stocks move together over time. The price spread of a pair of cointegrated stocks doesn't change drastically over time. While many stocks are correlated, a pairs trading strategy profits off of stocks that are both correlated and cointegrated. When the prices of the two stocks diverge, a pairs trading strategy buys the lower-priced asset and shorts the higher-priced asset. An investor generates a profit when the prices revert to their means. Investors profit if the prices converge, but lose money if they diverge further. Do and Doff (2010) outline the key premise of market inefficiency in this strategy. In an inefficient market, irrational trading will lead to occasional price divergences of similar securities. This creates an arbitrage opportunity.

Advantages of a pairs trading algorithm include market neutrality and hedged risk. By taking a long and short position in a pair of stocks, pairs trading focuses on the relative performance of stocks. Thus, there is little or no net exposure to the direction of the overall stock market. Our group wanted a low-risk investment strategy and believed we could apply this strategy across multiple sectors, further lowering our risk and diversifying our portfolio.

There are several psychological and economic rationales behind pairs trading. Jegadeesh and Titman (1995) suggest that individual investors tend to overreact to company-specific information shocks, rather than trade on fundamental common factors. This creates an opportunity for contrarian profits. Another rationale is the Law of One Price, which suggests that the prices should be identical for a pair of assets that have matching payoffs. Gatev Goetzmann and Rowenhorst (2006) argued that the profits from pairs trading are compensation for enforcing the Law of One Price.

Andrade di Pierto and Seasholes (2005) expanded on Jegadeesh and Titman (1995) by looking at the role of uninformed trading demand shocks in the profitability of pairs trading. As in traditional asset pricing models, we assume that stock returns are composed of loadings on systematic risk factors plus an idiosyncratic component. Within a large enough observation period, stocks that have historically moved together can be thought of as having similar factor loadings. If we assume factor loadings remain constant in the future, the stock returns should continue to move together. However, they argue that temporary divergence in stock prices result from idiosyncratic shocks which in turn lead to persistent price differences. When uninformed investors trade based on firm-specific information, they create demand in the stock market that is uncorrelated with asset fundamentals. Informed investors accommodate these demands, but require compensation because they are risk-averse. Thus, uninformed buying is followed by a rise in prices. After the demand shock, prices mean-revert back to fundamental levels. Likewise, uninformed selling is followed by a fall in prices and similar mean-reversion.

Backtest

Relating the Code to our Strategy

Our code looks at each Morningstar sector and identifies 1 pair of stocks per sector from running cointegration tests over price data of the ten largest market-capitalization companies in each sector in the period 2006-2010. It then selects the most cointegrated pair for each sector.

This is crucial to a pairs trading algorithm because we profit off of mean-reversion, i.e. we want a highly cointegrated pair of stocks. In a sense, the more cointegrated the pair of stocks is, the less risky the arbitrage opportunity is. This is done in the Jupyter Notebook before going into Quantopian.

In our algorithm, we use statistical tests to identify stationarity over a preceding number of days. The goal is to combine the tests of correlation and cointegration to find a stationary time series consisting of a linear combination of a pair of stocks. We use tests like the Augmented Dickey-Fuller (ADF) unit root test, the Half Life test from the Ornstein-Uhlenbeck process, and the Hurst Exponent. The ADF test looks at whether or not a security is mean-reverting, the Half Life test looks at how long the security takes to mean-revert, and the Hurst Exponent looks at how often the security mean-reverts. Once complete all of these statistical tests, we need to calculate the hedge ratio. The hedge ratio is used to determine relative stock prices. We use Ordinary Least Squares (OLS) to calculate the hedge ratio because it can take into account past prices. The formula then for getting the spread is Price A - hedge * Price B.

Every day, we run our statistical tests and see if they all passed. If all of the tests say that this pair creates a mean-reverting stationary time series, then we will possibly want to execute a trade. However, if one of the tests fails, we either close our position (if we have an open position), or we skip this pair today. Then, we calculate the Z-score for our spread, which tells us how many standard deviations away the current price is from the mean price over some look back window. We then check to see if the spread has increased or decreased. If the spread widens, we want to short the winner and buy the loser. If the spread narrows, we want to short the loser and buy the winner.