Statistical Arbitrage and High Frequency Trading Simulator

Team: Winnie Cheng (wwcheng@gmail.com) and Tirto Adji (tirto_adji@yahoo.com)

Overview:

Statistical Arbitrage is a set of trading strategies that make use of data mining techniques and observed statistical properties of groups of stocks to forecast future movements. This project aims at developing an R package and associated visualization tools for traders to analyze the effectiveness of their strategies. The package will also include a small set of sample strategies that the open-source community can expand on.

We will focus on the S&P 500 dataset for this package. Daily stock prices of S&P500 component stocks and their characteristics will be obtained from:

- http://www.standardandpoors.com/indices/sp-500/en/us/?indexId=spusa-500-usduf--p-us-l--
- http://pages.swcp.com/stocks/

The fundamental principle behind statistical arbitrage is in 1) identifying stocks that are correlated, 2) exploiting the mean-reverting behavior to expose trading opportunities. We will provide 2 simple strategies for demonstrating 1):

- Correlation by industry
- Correlation of past 6-months daily returns

For example, for two stocks A and B that are highly correlated, if stock A rises and stock B falls, and the difference is 'wide enough', the mean-reverting principle suggests that one should long (buy) B and short (sell) A, since the two stocks will 'revert' to the same mean later on. The package will explore the above two correlation strategy as well as providing a flexible simulator that can evaluate other strategies.

The Simulator will take in an algorithm and the S&P stock market data to compute the return on the entire portfolio over time. We intend to have plotting functions in R as well as interfacing to Google Visualization API to enable 'playback' of the trading strategy, giving traders greater insight into how their strategy played out in different time intervals.

The initial architectural diagram of our R package is presented below.

The intended use case is for one to select or write a trading algorithm. These trading algorithms typically involve a calibration or training stage in which they run over historical behavior to calculate features such as correlation coefficients. Then the algorithm is fed to the trade simulator module. This module runs the algorithm on market data to assess the return of the strategy. The output is presented as day-to-day portfolio returns which can be visualized in various ways, in R plots as well as via interactive charting using the Google API.

We intend to divide the work into two R packages: Strategy Simulator, and Portfolio and Market Visualizer. Both of these packages will make use of stock prices and info data. The objective of the Strategy Simulator is to use allowing the running of different trading algorithms on a time-series of stock data. It outputs a set of trading decisions based on these algorithms. The Portfolio and Market Visualizer provides algorithm designer feedback on how their overall portfolio is performing as well as how individual stocks have performed over the simulated time period.