**CS - 1 : Quantitative Analysis and Modeling for S&P 500**

**Aakriti Sinha**

**9th October 2020**

**Introduction**

The meteoric increase in compute power and advances in Machine Learning have given rise to a variety of use-cases for mechanical/algorithmic trading. Quantitative funds across the world use a plethora of techniques to forecast market prices, volumes and general market behaviour. S&P 500 is one of the world’s leading benchmark indices consisting of 500 publicly listed companies. This study will be restricted to data from these companies’ price-volume data (as traded on the New York Stock Exchange).

**Objectives**

There are 3 main objectives -

**1. Volatility Index** - Out of all the 500 stocks in the dataset, establish a weekly volatility index which ranks stocks on the basis of intraday price movements. (Weekly volatility Index implies that it is to be calculated on a weekly time frame and both intraday as well as weekly change in price needs to be used in calculating volatility)

a. Give an exploratory analysis on any one stock describing it’s key statistical tendencies.

b. The index should rank the stocks from most to least volatile in the selected time frame.

c. The output needs to be grouped weekly showing the Top 10 Most and Least Volatile stocks. Both your code and output will be evaluated.

**2. Pair Trading** - The concept of pair trading suggests that there are stocks whose prices move together (could have an inverse relationship). The objective is to identify the 5 strongest pairs for every year in the dataset (eg. 5 strongest pairs for 2014, 2015 and so on).

**3. Binary Classification** - Given a stock and it’s data, you have to predict whether it will close lower than it opened (red) or higher than it opened (green)

**Approach**

OBJECTIVE 1: The approach for the first objective is statistical in nature. Volatility can be measured by the standard deviation of returns over a chosen period of time. Historical Volatility (HV) is calculated as:

HV = (Std dev of Price Returns) \* (Square root (T))

where T is the number of periods in chosen time frame. Price Returns can be calculated as the natural logarithm of price change over a single period. For weekly VIX, price returns are calculated using daily price change. The number of periods in a week is 5 (counting only weekdays as the stock market is closed on weekends).

Once the volatility index is prepared, the next step is to rank stocks. The ranking needs to be done on a weekly basis to identify the top ten most and least volatile stocks each week. We use the average volatility of a stock on the last day of week – Friday – to calculate the rank.

The final section of the first objective is related to the statistical exploration of any one stock’s performance in the market over the given time period. We have chosen to study Starbucks Corp. The Starbucks stock is examined through various market measures:

* Trend of closing price
* Trend of volume
* Frequency distribution of opening price, closing price, high , low and volume
* Daily logarithmic price returns
  + Log(closei – closei-1)
* Daily return or Percentage change in daily closing price
  + (closei – closei-1)/closei-1
* Cumulative daily return
  + Cumulative of daily logarithmic price returns
* Moving average for 10, 20 and 50 days
* Auto-correlation with a lag time of 5 days
* Weekly Volatility Index (as calculated above)

All these parameters were also plotted on a graph to enable visual analysis too.

Risk analysis was subsequently conducted for Starbuck Corp.

* Expected return rate
  + Average daily return for the time period
* Total return rate gained from the stock
  + (Daily Return last day of time period – Daily Return first day of time period) / Daily Return first day of time period)
* Total risk
  + Standard deviation in daily returns
* Sharpe ratio
  + ((Average return rate – Risk-free return rate) / standard deviation of returns) \* sqrt(252)
  + Risk-free rate is the lowest risk offered by any investment portfolio in the market. It is usually assumed to be 0.01.
  + The ratio is annualized by multiplying by square root of average number of business days a year.
  + The standard deviation represents the volatility of the stock i.e. how much the return deviate from expected value. The returns could be higher or lower than expected value.
  + The higher the Sharpe ratio, more stable or less volatile a stock is said to be. A value above 1 is said to be good. If a stock has Sharpe ratio above 2, it is believed to be very good with low volatility. Excellent stocks reach values above 3. However, a negative value denotes high volatility, which shows that the stock is not reliable.
* Sortino Ratio
  + ((Average return rate – Risk-free return rate) / standard deviation of downside returns) \* sqrt(252)
  + While similar to Sharpe ratio, sortino ratio is often studied to understand the downside risk. Deviations in return on the upper side of the expected values are welcomed by investors and is considered good risk. Downside risk denotes loss in value. Thus, sortino ratio penalizes only the harmful downside risk.
  + The higher the Sortino ratio, the less risky the stock is. A value above 1 is said to be good. If a stock has Sharpe ratio above 2, it is believed to be very good with low downside risk. Excellent stocks reach values above 3. However, a negative value denotes high risk.

OBJECTIVE 2: The second objective required to identify the strongest pairs of stocks of each year. Pairs trading is widely followed as an investing or trading strategy. It revolves around the concept to simultaneous movement of a pair of stocks. Certain stocks are affected almost identically by market conditions. These often belong to the same portfolio or industries. Traders look for deviations in their movements i.e. difference in their stock prices. When deviation in trends of a pair is large, there is opportunity to trade.

Pair trading analysis involves looking for these very deviations. Machine learning techniques can help find patterns that people may miss. In our case study, we are using these techniques to identify such pairs of stocks that generally move together in the market. A huge correlation matrix is built with each pair of stocks trading in a year. Each pair of stock is measured with the Pearson’s correlation coefficient. Each pair is thus sorted in descending order to identify the tem most and least correlated pair.