Applications of Quantitative Analysis in Stock Trading

Omar Abdelkader, Ben Cannon, Ryan Elder, Erich Meissner

***Abstract*—Quantitative research has become universal at all investment banks, hedge fund, and private equity firms. Large financial institutions are hiring mathematical and statistical doctorates now more than ever. The potential returns of algorithms built off probability concepts are extremely promising, and several hedge funds have posted astronomical results that are too good to ignore. However, catastrophic failures in these algorithms in financial engineering, such as the story of *LTCN*, can occur just as often. Due to advancements in web applications for trading and stock analysis, we can test our theories on past performance. We will test our ideas with the popular quantitative analysis tool, *Quantopian*, to see how our engineering-probability-course-based Python script performs against the benchmark *S&P 500* index.**

***Index Terms*—Moving average, Bollinger Band, trade volume, market cap, S&P 500.**

I. Introduction

In this project, we analyze various simple statistical models used to trade securities in the stock market. We back-test each of these strategies against the S&P 500, dating back to 2006. This time-frame was selected deliberately to include the period known as *The Great Recession* (Dec. 2007 — June 2009) in order to fully analyze the robustness of each model under test. In addition, we choose to only trade mid and large cap stocks, or stocks with a market cap of $2 billion and above. A stocks market capitalization, or *market cap*, is defined as the number of outstanding shares, multiplied by its current price. Mid and large cap stocks allow us to minimize our risk, while maximizing our growth potential.

We consider two primary trading strategies: The first strategy involves buying and selling based on the moving average of a stocks closing price. The second strategy involves trades made based on the relation between the stocks closing price and its *Bollinger Band*. By definition, the closing price of a stock is the price when the market closes at 4:00PM EST. We choose to use the closing price because it disregards any fluctuations made in after hours trading.

II. Moving Averages

A. Relevant Theory

Lets assume we wish to find the expected value of some discrete random variable, X. Normally, to calculate the expected value of the discrete random variable X, we would perform a summation of all the different possible values of X weighted by their corresponding probability mass function (PMF) values.

$$E[X] = \sum_1^N x_i P_x(x)$$

But in some situations we do not have access to the variables PMF. However lets say we do have historical data on different values the variable has taken on in the past and how often these values have occurred. We can approximate the expected value of X by taking an average. If we have N distinct values of X and the i^{th} value of X, say X_i, occurred k_i times, we say that the average value of X equals:

$$\overline{X} = \frac{x_1k_1 + x_2k_2 + \cdots + x_Nk_N}{k}$$

Where $k = k_1 + k_2 + ... + k_N$. As k grows very large (and we look at probabilities as relative frequencies) it makes sense that the fraction $\frac{k_i}{k}$ would approach $P_x(x_i)$ and the average we took of X would approach *E[X]* [1].

B. Stock Trading with the Moving Average

The simple moving average is a commonly used metric in stock trading. In our algorithm, we calculated the arithmetic mean of a given security by summing the closing prices over a preselected window size and dividing that total by the window size itself, similar to the method discussed in the previous section. A short-term (smaller window size) moving average crossing above, or exceeding, a long-term (larger window size) moving average, is typically indicative of an uptrend in a stocks price. Conversely, when a long-term moving average crosses above a short term-moving average is indicative of a pullback, or downtrend in price. Short term moving averages are better indicators of a stocks momentum trajectory because smaller window sizes will only capture the most recent data points. While the large window will also capture those same data points, it will also include values that are further from the current close price.

Two commonly used trading strategies that use the simple moving average are known as the golden cross and death cross. The golden cross happens when the 50-day moving average passes through 200-day moving average. The death cross, on the other hand, occurs when the 200-day passes the 50-day [4]. Historically, the case golden/death cross is compelling, however, we felt that using even smaller window sizes, specifically the 20 and 50-day, would give us a better opportunity at securing smaller initial gains in order to grow our capital faster. This is because the 20-50 and 50-20 day crosses occur more often than the golden/death crosses.

In the event that multiple securities satisfy the buy condition, a tie breaker will be decided upon using the securitys trade volume. Since we are looking to confirm an uptrend in price, we will look for high volume from buyers [5]. By calculating the percent difference in volume over a given window, we can fairly assign numerical weights to each security and buy in accordance with the rest of the market. This strategy also works in reverse in the event that multiple securities satisfy the sell condition.

III. Bollinger Bands

A. Relevant Theory

The second part of our project relies on the concept of standard deviation. The standard deviation of a random variable is derived from the variance of that random variable. The variance of a random variable shows how much the value of the variable varies around its expected value. Lets say X is a random variable. The variance of X equals:

$$Var(X) = E[(X - E[X])^2]$$

This can be viewed as our expectation of how much X will differ from its expected value squared. For our situation, where we do not know the explicit PMF of our random variables, we can approximate the variance by taking an average of $(X - \overline{X})^2$:

$$\sum_1^N \frac{k_i(x_i - \overline{X})^2}{k}$$

Where X_i, k_i, k, and N are as defined in the section discussing moving averages. Once we have an approximation for the variance, we can then find the standard deviation, $S(X) = \sqrt{(Var(X))}$.

The standard deviation of X can also be looked at as a measure of how the values of X vary from the expected value (or in this case the average), but can be easier to understand than the variance since it has the same units as X. For example, lets say we find X has an average of 4 and a standard deviation of 2. We would expect typical values of X to vary between $\overline{X} \pm S(X) = 4 \pm 2$, or 2 and 6 [1].

B. Stock Trading with the Bollinger Bands

The Bollinger band trading strategy can be seen as an extension of the simple moving average pattern. We select a window size, say 20 days, and compute the standard deviation of closing prices of that sample. Standard deviation, in the context of stock prices, can be viewed as a measure of volatility within the sample. The intuition behind trading using the Bollinger bands is in mean reversion. Whenever market sentiment drives the price of a stock outside two standard deviations of the moving average, the efficiencies of the market drive the share price back to its mean, or moving average [3][6].

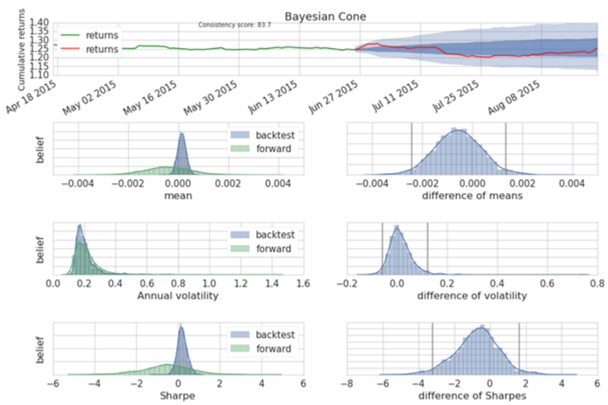
It is worth noting that the price of a security is not normally distributed and techniques such as the central limit theorem are unreliable. This is because it is incorrect to assume that each price is independent from the others in the sample. In fact, each price will be very closely related to the ones closest to it in time. Therefore, we cannot claim that 95% of the data will be encompassed by the bands, as you would in a normal distribution, however studies have shown that 88% of a securitys prices will remain within the bands [2]. Therefore, we can still safely assume that the price will revert to the mean.

Knowing this, we plan to buy securities that fall below two standard deviations of the moving average, and sell when the price jumps above two standard deviations of the moving average. Similarly to the simple moving average strategy, ties will be broken using trade volume.

IV. Bayesian Analysis & Simulations

Over the course of this project, we noticed that the back-test might fit well with past data but will fail on unseen data. This case is similar to the one Professor Martins warned us about. He told us a story of a colleague of his who applied probability theorem to the markets, and the algorithm failed horribly due to unpredictable irregularities in stock volatility. We found it promising to apply a Bayesian estimation and prediction of possible future returns. The Bayesian model takes the time series of past daily returns using an algorithm as input. The end result is a simulation of possible future returns.

For example, comparing the actual performance of a trading algorithm on unseen market data with the predictions generated by our model can inform us whether the algorithm is behaving as expected based on its back-test or whether it is over-fit to only work well on past data. The Bayesian model enables this stock market vision. Professor Nuno Martins warned us of employing probability to the markets by telling us the story of *LTCN* described above. Such probability-based algorithms, similar to the ones that colleagues of Professor Martins may have developed, have the best back-test results but they may not necessarily have the best performance in live trading. An example of such an algorithm can be seen in the picture below. As you can see, the live trading results of the algorithm are completely out of our prediction area, and the algorithm is performing worse than our predictions. These predictions are generated by fitting a linear line through the cumulative back-test returns. We then assume that this linear trend continuous going forward. As we have more uncertainty about events further in the future, the linear cone is widening assuming returns are normally distributed with a variance estimated from the back-test data. This is certainly not the best way to generate predictions as it has a couple of strong assumptions like normality of returns and that we can confidently estimate the variance accurately based on limited back-test data. Below we show that we can improve these cone-shaped predictions using Bayesian models to predict the future returns[7].



The overall output of the Bayesian model, back tested during the stock market crash of 2007 - 2009, is outlined below:

Our team also found it interesting to back-test against later data after the markets had rebounded. We found this Bayesian model, which manipulated Bollinger band and moving average data, to perform no better than the benchmark S&P 500. However, we set out to lower losses and exposure during the housing crisis and resulting crash. At the end of the day, the best decision would have been to limit exposure completely and pull all assets out of the markets. The results outside of the time period of the market crash are posted below from 2013 to 2017:

V. Conclusions and Suggestions for Further Research

In conclusion, we found that there is no 'silver bullet' trading strategy for making a killing in the stock market. That being said, prudent strategies such as diversifying your portfolio or investing in ETFs along with monitoring basic indicators including, but not limited to, various moving averages and the Bollinger bands will play well and generate comparable returns to the market.

Acknowledgment

We would like to thank Professor Nuno Martins and TA Sai Saketh Rambhatla for their guidance and support throughout the course this semester.