**Capstone Final Project: Model Selection and Justification**

Objective: The goal of this project is to explore different technical analysis strategies to evaluate their performance against a passive buy and hold investment strategy. By incorporating a number of investment strategies and machine learning optimization techniques such as hypertuning this should give the foundation of building my own investment strategy.

The project's problem: There are currently so many different ways to use technical analysis techniques and it can be confusing to understand which performs best. It has been traditional to use algorithmic trading but with the increase in power of computing and the use of machine learning/AI the power in optimizing investment strategies are not much further in depth than some trading rule. That is where that project is going to try and bridge the gap between traditional static strategies with dynamic and intelligent strategies.

The goal of this is to explore a variety of different technical analysis techniques which are

* Bollinger bands
* Relative Strength Index
* Moving Average Convergence Divergence
* Ichimoku Clouds

**How do different technical analysis techniques combine to create a more powerful investment strategy?**

Each of these will be performed on Sector ETF’s during different time periods. This is done so that the model can be built off of a variety of performing sectors which is meant to be most generalized to prevent overfitting. Another advantage is that you are able to see which sectors are best performing during different invest strategies and in different periods.

For example you might have a portfolio with 10 stocks

Stock A might be AAPL

AAPL is a part of the technology (XLK) sector and if this sector has performed particularly well during an expansion then the strategy should hopefully be generalized to perform better when the stock is moving in a certain direction (suggestive of a business cycle). The stocks will have buy/sell signals from a variety of different strategies as listed above and this can also be incorporated into giving a specific weight of investment power. All of this combined should be able to create a powerful investment strategy.

By combining a number of these strategies whilst optimizing returns using a variety of machine learning techniques a final model should be created.

By utilizing the foundations of stochastic modeling where different starts dates are going to be used to prevent overfitting a distribution of returns is going to deliver an expected return value.

**Model 1. Buy and Hold Passive Strategy**

The buy-and-hold strategy is going to be used as the baseline investment strategy. It is a passive and low-stress strategy that any successful model should be able to outperform.

Stochastic Modeling Methodology for buy and hold strategy:

The following is going to occur for each Sector ETF within each period:

1. Choose 1000 start dates

2. Purchase a single stock

3. Hold onto the stock

4. Sell the stock at the end of the 90 day period

5. Record the ROI as a list in a nested dictionary

6. Repeat 2-5 for each 1000 iterations

The following is an example of a 90-day buy-and-hold distribution of the ROI of the healthcare sector ETF, XLV during a trough

A graph of a growing graph

Description automatically generated with medium confidence

Observe that the distribution is almost normal which suggests that for XLV during a trough you can expect almost a -5% return with a 5% return being extremely unlikely, dependent on the start date of investing.

Sector Performance

The following is the table of returns using a buy-and-hold strategy during different macroeconomic cycles. These returns took place over 90 days.

A table with numbers and symbols

Description automatically generated

The average result of this investment period combining all of the sector ETFs is:

A black background with white text

Description automatically generated

Observe that the trough period performs the worst with sectors such as financials and real estate returning as low as -22% and 18% respectively. Meanwhile during a peak the financial returns 10% and technology a 12% this clearly shows that this investment strategy is heavily dependent on the time of investment. If you are able to understand the macroenvironment and only invest during an expansion and a peak ,expect a 6% return during a 90 day period however this is extremely hard to do. Making a model that tries to perform better when the overall economy decreases is going to be the challenge!

Dashboard Creation

See video link to dashboard here - <https://www.youtube.com/watch?v=adQphpCyI4w>

The dashboard as shown below has a dropdown for time period and stock to show the distribution function of returns.

A screenshot of a graph

Description automatically generated

The dashboard would be extremely useful for individuals who are looking to invest into a stock. Again using AAPL as a stock you could look at the XLK technology sector ETF and see the likely return based on the macroeconomic cycle you are in.

The standard deviation is a metric used to measure risk and volatility. It indicates the dispersion from the mean, which can be seen in the distribution of results. As the standard deviation increases, there is a higher risk because the returns have a wider spread. When the standard deviation is lower, the expected return is much more precise.

A table with numbers on it

Description automatically generated

The average standard deviation across the buy and hold strategy is

A black background with white text

Description automatically generated

**Model 2. Bollinger Bands**

Bollinger bands use momentum indicators to create a channel for price of stock to move between, if it falls below a buy signal is triggered and if it rises above a sell signal is indicated. This can be used as a technique to invest in stocks.

The parameters that go into bollinger bands are the rolling average period and the standard deviation or the confidence interval that is associated.

Model 2. Baseline Bollinger Band Model

The baseline model is what John Bollinger himself suggested which was to use a 20 day buy with a confidence interval of 95% or about 2 standard deviations above and below the 20 day rolling average.

Baseline Bollinger Bands Stochastic Modeling Methodology:

1. Choose 1000 start dates

2. Iterate through each ticker of each day

3. Perform action

* Buy: purchase 20% of account balance (starts at $100)
* Sell: Sell 20% of current stock holding
* Hold: Go onto the next ticker/day

4. Update the current holdings through multiplying stock price by number of shares owned

5. Continue for the 90 day investment period

6. Record ROI after the 90 day period

7. Store ROI for each stock in each period as a nested dictionary

The below is the return of a healthcare sector stock (XLV) during a trough time period

A graph of a graph

Description automatically generated with medium confidence

The following is a return on investment for each stock in each period

A table of numbers on a black background

Description automatically generated

The following is the overall return for each period using the baseline bollinger band model, observe that during a trough it performs quite well.

A screenshot of a computer screen

Description automatically generated

Model 3. Parameter Optimized Bollinger Band Model

The rolling average [10,15,20,25,30]

The confidence interval would be [0.85,0.90,0.95]

The grid search is an optimization technique which will brute force test each different combination to try and find the highest ROI

A table with numbers and letters

Description automatically generated

The highest average ROI is a rolling window of 30 and a confidence interval of 0.90 which means that the higher returns take place where there are more buy and sell signals that occur.

Perform stochastic modeling with optimized values:

A table of numbers on a black background

Description automatically generated

There is a greater improvement to the model with some sectors such as financial seeing a 10% return on a 90 day investment.

The average return across the different sectors

A black background with white text

Description automatically generated

It is clear that Bollinger Bands do well during a trough and the optimized parameters boost that to be an even stronger indication.

The following is the healthcare sector ETF return during a trough time period

A graph of a graph showing the value of a company

Description automatically generated with medium confidence

A graph with red green and blue lines

Description automatically generated

Observe that the investment period takes longer to begin, so you are only actually getting 60 days of availability to trade. However during that time there are more trades that take place because the confidence level is only at 90%.I believe that there can be a multiplier effect so that during these cluster periods of buy and sells there is more investment with each incremental signal.

Evaluation of base model against benchmarks (buy and hold strategy)

The base model showed that during the trough time period the success rate was far better, over 10% better with some industries showing returns of 25% on average over 1000 iterations which is quite impressive for a 90 day investment period.

An investigation can be done to understand why the trough and contraction showd better performance, it is likely because of the pace of the improvement of peak and expansion outpaces the profits made from an investment strategy such as bollinger bands.

A table of numbers on a black background

Description automatically generated

This shows the average return

A black background with white text

Description automatically generated

Evaluation of optimized model against benchmark (buy and hold strategy)

By using the optimized parameters as seen above in the model creation it performs well but it performs even better when you compare it to the benchmark.

A table with numbers and symbols

Description automatically generated

You can again see that during the trough period the stocks all perform well with Financial, Real Estate and Industrial Sectors all performing 20-33% better than passive investment strategies. Meanwhile it again performs about 5% worst during peaks and expansions.

A black background with white text

Description automatically generated

It performs 14% better during a trough and overall you can see in some industries such as Finance it performs 33% better than a traditional buy and hold strategy. It is clear that the performance during peak and expansion periods (where economic activity is higher) the technical analysis strategy cannot keep up with the performance of the overall market.

Model 4. Optimizing investment amounts Bollinger Band Model

The amount that is invested for each buy/sell signal may greatly impact the profits. If you take bigger investments when the market is showing a higher likelihood of a rebound then you are likely to return a higher profit. Using grid search again, the bollinger band model is going to use the optimized 30 day rolling average period with a 90% confidence interval for investing.

Another reason this is important because as you can see below, the end of the investment period most of the values of returns are close to 0, this is because the number of sell signals are about twice as high as the buy signals.

A graph of a graph

Description automatically generated with medium confidence

Where there was 25 buy signals and 57 sell signals so it is likely that whenever a stock is bought it is not held on for long enough to see returns.

To try to over come, using different percentage multipliers to the amount to invest can be a good way to overcome this. Perhaps during an expansion, a higher amount to buy is going to perform better as the stock continues to rise with a lower sell rate.

The following will the parameters tested using the brute force grid search method

Amount\_to\_buy = [0.10,0.20,0.25,0.30]

Amount\_to\_sell = [0.10,0.20,0.25,0.30]

Gridseach results

A table with numbers on it

Description automatically generated

The higher ROI are all quite similar and the amount invested does not seem to make too large of a difference when looking at the average ROI. However, a slightly higher buy amount compared to sell amount is going to return the highest returns, with the combination of having a lower buy amount and a higher sell amount returning less favorable returns. This is likely because there are too many sell signals with too high of a sell amount which does not allow for the value of shares to increase.

Performing stochastic modeling using amount to buy as 0.30 and amount to sell as 0.20 with a 0.90 confidence interval and a 30 day rolling average

A table of numbers on a black background

Description automatically generated

The following were the mean returns on the investment

A black background with white text

Description automatically generated

**RSI Model**

The goal of the RSI is to create a dataframe of signals based on the thresholds as explained above. This can then be used to analyze the performance of incoprorating RSI signals in comparison to a passive buy and hold strategy.

Use n days for the time period where you are looking at average gain/average loss. It is typical to use 14 days as recommended by J. Welles Wilder but longer and shorter time periods are also utilized.

Another parameter that can be looked at is the minimum RSI value before a buy/sell signal is created.

**Methodology**

1. Difference the data

2. Track average gain/loss where gain is the positive movement from day to day and loss is the negative movement

3. Create the RS via the average gain/average loss

4. Create the RSI value by 100 - (100/(1+RS))

Model 5. RSI Base Model

* 14 day investment period
* 0.20 buy/sell amount
* 0.30 as the minimum to trigger a buy signal
* 0.70 to trigger a sell signal

A table with numbers and symbols

Description automatically generated

The average roi for different time periods using the RSI signals.

A black background with white text

Description automatically generated

**A graph of stock market data

Description automatically generated with medium confidence**

**A graph showing a graph of a number of people

Description automatically generated with medium confidence**

**MACD**

The MACD technical analysis will create the following parameters:

- Short Exponential Moving Average (12-day)

- Long Exponential Moving Average (26 days)

- MACD Line: The difference between Short EMA and Long EMA

- Signal Line EMA of the MACD line

By utilizing an exponential moving average it gives a greater weight to the more recent values, this makes it more sensitive to recent changes in price. The MACD line measures the convergence or divergence between fast and slow moving averages. When the MACD line is positive (above 0), the short term trend is stronger than the long term trend which suggests an upwards trend. Where as when the MACD line is negative (below 0) the short term is weaker than the long term trend which suggests downwards momentum.

The signal line is used as the buy and sell signal by smoothing the MACD line using a 9 day ema period. When the MACD line goes above the signal line it indicates a buy signal and the MACD line goes below the signal line it indicates a sell signal.

**A table with numbers on it

Description automatically generated**

**A black screen with white text

Description automatically generated**

**A graph of a stock market

Description automatically generated with medium confidence**

**Ichimoku Cloud**

**A table of numbers on a black background

Description automatically generated**

**A black background with white text

Description automatically generated**

**A graph with lines and numbers

Description automatically generated**