THE ROBO-ADVISOR AND SMART INVESTMENT FUND

ELEC7080 Algorithm Trading and High Frequency Trading

ABSTRACT

The Robo-advisor and Smart Investment Fund is a system that focuses on the analysis of the Standard & Poor's 500 (S&P 500) index. Using several indicators to generate market view, stock selection strategy, trading signal, and long-short strategy. Offering information and flexibility in trading to investors.

TEAM MEMBERS

WONG Nga Yan, Dorothy

Park Seo Hye, Sarah

Ho Kam Tim, Steven

Li Zhilin, Camille

Table of Contents

1. Introduction & Parameters Selection

1.1 Introduction About the System

Getting wealthy is a dream to many, and making investment is one of the avenues. While investing, a lot of people wish to make outsized returns, yet at the same time they would strive to maintain stable performance in the long run and minimize potential drawdown.

However, many investors might not have time to get into extensive and in-depth research in each of the asset classes, sectors, companies specifics and even stay tune for asset movement during the day. To help these investors, we are proud to introduce The Robo-advisor and Smart Investment Fund.

Traditionally, investment have been focusing on both fundamental and technical analysis to achieve a portfolio to out-performance benchmark. Our model will be using a hybrid of both kind of parameters, to find out their correlation to each of S&P500 constituents on the following day.

Nobody has a crystal ball to tell when a significant event is going to shock the financial market. But that does not imply there is no way to deal with market volatility. Our proposed strategy is to long select stocks which our model predicted to outperform and short those underperform. By doing so, investors can achieve a low-beta near market-neutral portfolio which focuses on capturing alpha, while investors can also enhance the portfolio performance by adopting various beta strategies, or even other asset classes which fit their investment need and thesis.

1.2 Discussion on Parameters in the Prediction Model

1.2.1 VIX

The Cboe Volatility Index (VIX) derived from the prices of SPX index options with near-term expiration dates, generates a 30-day forward projection of volatility. It is an important index in the world of trading and investment because it provides a quantifiable measure of market risk and investors' sentiments. Throughout the rally in the stock market in 2020-2021, VIX has been falling, and hitting new lows each time the SPX hits a new high. While negative volatility-spot correlation is not unusual, it is a break from the trend of the last three years, in which new highs were often associated with nervousness about valuation levels.

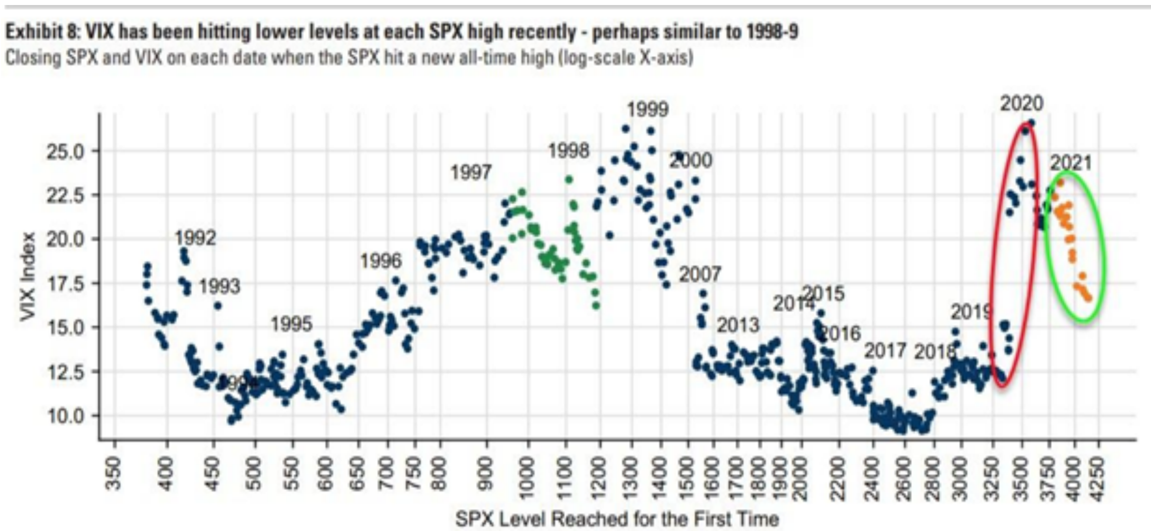


Figure 1: VIX against SPX Level Reached for the First Time

1.2.2 Treasury Yield

10-year Treasury Yields are a benchmark of risk-free rate and more importantly expectation of inflation in near term. Using the money supply as a proxy, we can compare the money supply changes to inflation. The chart below advances M2 by 9-months as compared to CPI and the Fed Funds rate. If the historical correlation holds, the Federal Reserve will have pressure to taper monetary policy and hiking interest rates. And the last time the Fed started discussing similar policy changes was in 2017 (Figure 2), which led to a sharp pull back in the stock market.

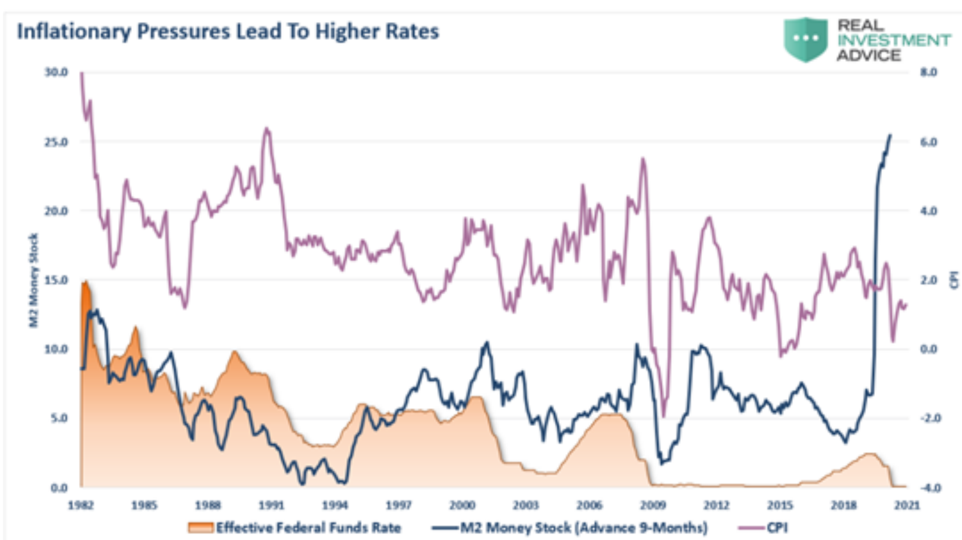


Figure 2: Inflation Pressure led to Higher Rates

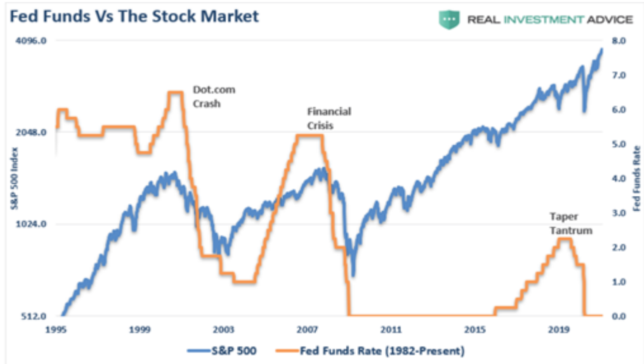


Figure 3: Fed Fund Rate vs S&P500

1.2.3 Factors ETFs

For the Momentum factor indicator, we are using iShares MSCI USA Momentum Factor ETF, which is one of the most popular ETF issued by Blackrock which tracks a Net Asset Value (NAV) of $13 billion. This ETF consists of a basket of US stocks which have been outperforming in the market within the last rebalance period. And on the other side, we also include iShares MSCI USA Value Factor ETF, this Value factor indicator is a proxy of recent underperforming stocks with a Net asset value of $22 billion. Below exhibits a small comparison of the 2 ETFs, we can see the P/E and P/B ratio of the 2 funds are very different which also suggests they are capturing a different factor or drivers in market trend.

Figure 4: iShares MSCI USA Momentum Factor ETF

Figure 5: iShares MSCI USA Value Factor ETF

1.2.4 DXY

DXY is an index of the value of the United States dollar relative to a basket of foreign currencies. When USD appreciates against other currencies, since the price of commodities (oil, copper, iron ore) are quoted in USD, various raw material prices are effectively lower (in USD terms) and that will be beneficial to economic growth and the general economy.

Figure 6: DXY Index vs US economy growth expectation

2. Pre-trade Analysis

This section will introduce the details and procedures on pre-trade analysis. To study from different dimensions in market view generation and trading opportunities identification, both structured data and non-structured data are considered in the data acquisition and collection process. To facilitate computation speed and allow easy retrieval during prediction model and trading strategies generation processes, the data collected will be stored in our SQL database and at the same time visualized into a web application. The URL for the demo web application is https://elec7080.herokuapp.com/, details and the UI layout of the website and its elements are shown in the Appendix. Generally, our Pre-trade analysis can be divided into six important aspects and will be discussed as the followings.

2.1 Macroeconomics Aspect

Major economic events and correlations of different world indexes are collected from the Federal Reserve Economic Data, Google Finance and Yahoo Finance. The economics events are combined into economic calendar data frame, which will include not only the actual figures, but also historical and prediction figures from the market to estimate the impact on volatility.

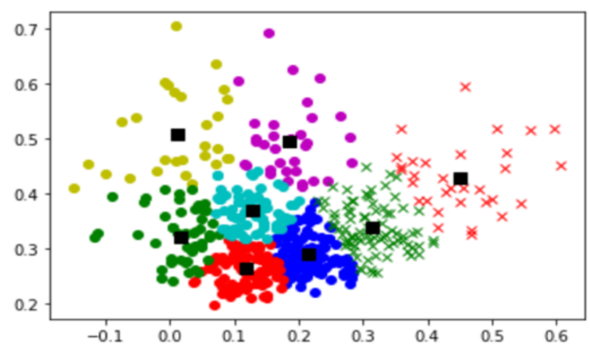
Using the daily return and volatility figures of different world indexes, a World Indexes Correlation Matrix can be formed, where 0 indicates not correlated and 1 indicates an exact match.

	previous	time	country	impact	event	actual	forecast
0	1.3%	2021/05/24 01:00:00	Singapore	Moderate Volatility Expected	CPI (YoY) (Apr)	2.1%	2.0%

Figure 7 World Indexes Correlation Matrix

2.2 Market Aspect

In this stage, S&P 500 stocks historical prices and details are acquired to calculate the percentage average return and volatilities. The data are served as inputs of Keras machine learning model to perform a K-means Clustering for the S&P 500 stocks. In particular, the 500 stocks are divided into 8 clusters based on x-axis: % average return and y-axis volatilities in a specified period. From the K-means clustering results, we can not only compare the performance of the 500 stocks, but also can rate and rank the stocks inside clusters and within an industry, as well as simulate a portfolio like the benchmark S&P 500 by combining the mean-valued stocks in each cluster.



	Stock	Cluster
0	A	1
1	AAL	6
2	AAP	3
3	AAPL	0
4	ABBV	1
5	ABC	4
6	ABMD	2
7	ABT	0
8	ACN	1

Figure 8 Clustering for the S&P 500 stocks

2.3 Technical Aspect

After looking into macroeconomics and market aspects, we will move on to analyze S&P 500 stocks individually. SMA/EMA, Bollinger Bands, MACD and RSI will be studied:

SMA/ EMA/ MACD: to compare different periods of moving average to study the market trend and momentum. Golden cross and death cross will be identified. The golden cross appears on a chart when a stock's short-term moving average crosses above its long-term moving average. The golden cross can be contrasted with a death cross indicating a bearish price movement.

Bollinger Bands: two standard deviations above and below the 20-day moving average. Widening of the bands shows increased the volatility and narrowing of the bands shows decreased volatility.

RSI: compares the size of up moves with that of down moves during a specified period (e.g. 14 days in our case), and indexes them to a scale of 0 to 100. Typically, security will become overbought when RSI is above 70 and over sold when the RSI is below 20.

2.4 Fundamental Aspect

Fundamental analysis focuses on predicting the intrinsic value of the stock (based on its revenues, profit, costs, capital structure, cash flows, and so forth), and analyses the stock price from the macroscopic and microscopic aspects to identify stocks that are not correctly priced by the market. Thus, Company metrics can then be compared with industry peers and competitors.

2.5 News Aspect

News from different sources and major medias are also one of our data acquisitions targets. Real time news is collected using python web scraping algorithms from investors.com, yahoo finance, Reuters and Bloomberg etc. Then, simple NLP sentiments analysis is performed using python Natural Language Toolkit (NLTK), an opensource Python library for Natural Language Processing. The stock news headlines are then classified into negative, neutral, and positive, each with scores from 0 to 1 to calculate the compound score. If the compound scores are closer to one, there is higher possibility that it is a positive news and vice versa.

	Date	Time	Headline	neg	neu	pos	compound
AAPL	2021-05-24	11:45AM	10 Best Dividend Stocks to Buy According to Billionaire Ken Fisher	0	0.7040	0.2960	0.6369

Figure 9 News from different sources and their polarities

2.6 Analysts/Price Makers Aspect

The final aspect that would be considered is the Analysts/ Market makers where the Analysts' ratings and recent insiders trades would be listed and recorded. Analysts' ratings will include the recommendations and forecasts from investment banks and private equities firms' famous analysts. While recent insiders' trades will record if there are any stock holdings movements from major shareholders, which are the price makers of the stock.

Analyst Ratings:

	Date	Level	Analyst	View	Predition
0	Apr-29-21	Upgrade	Goldman	Sell → Neutral	$83 → $130
1	Apr-06-21	Reiterated	Morgan Stanley	Overweight	$164 → $156

Recent Insider Trades:

	Trader	Relationship	Transaction	Cost	# Shares	Value ($)	# Shares Total	SEC Form 4
May 03	Adams Katherine L.	SVP, GC and Secretary	Sale	132.80	17000	2257631	328174	May 05 06:30 PM

Figure 10 Analysts' ratings towards stocks in the market

In addition to the above, Google Search trends for the stocks will be recorded. Interesting patterns are identified. Generally, upward Google Search trends relates to the volatility of the stock prices.

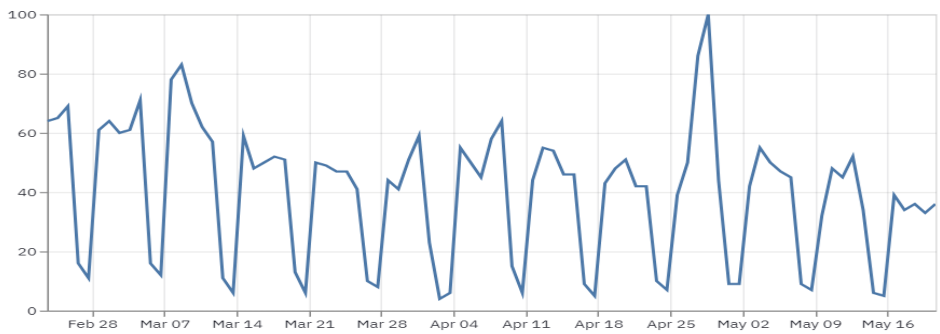


Figure 11 Google Search Trends

After studying and interpreting the outputs from the above six aspects, data are further consolidated and processed to form a comprehensive report. Instead of solely provide the figures for individual stocks, the report will be focused on the comparison with SPY as a whole/ another stock as the benchmark, which includes the cumulative returns, distribution of returns, Sharpe ratio, Maximum drawdown, monthly return and breakdowns, Key performance matrix comparison and so on.

As seen from above, both structured and non-structured data are collected during pre-trade analysis. These data can be integrated to generate trading signals and identify opportunities for arbitrage, especially for the later stage's prediction model and trading strategies formations. In case it is suggested that positive outlook is generated or predicted, one can enter the market from arbitrage and vice versa.

3. Prediction Model

3.1 Data Exploration & Data Split

Before building the prediction model, we delivered data exploration analysis. Since most of the raw data was not stationary, that might cause some problems when building the model.

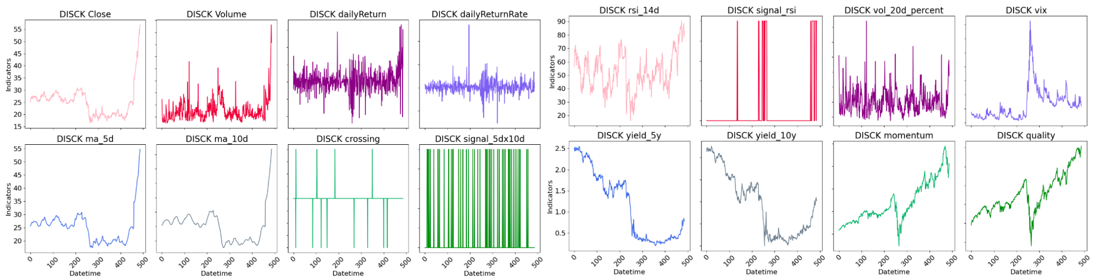


Figure 12 Trend of different indicators within a period

Thus, we implemented differentiate and percentage on the raw data. After that, the data was stationary.

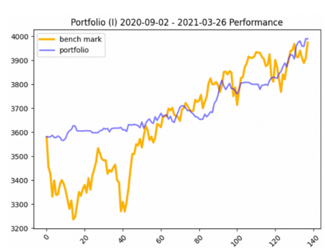
Then we split the data into three sets, the training, validation, and test set. The training set was used to train the model. The validation set was used to evaluate the model performance and provide information for selecting stocks as well as generating trading signals. The test set was used for back-testing.

3.2 Regression Model

The main idea of our long-short strategy is to use predicted return rate to select stocks and construct the strategy. Thus, the stock selection and the trading signals are based on the prediction model.

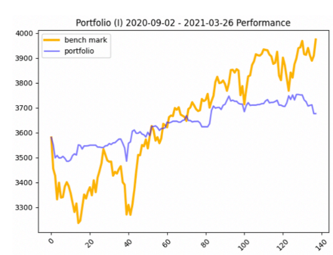
We used linear models like linear regression, robust regression, kernel regression, to find the relationships between the 8 significant attributes and next day's return rate. The kernel regression and the robust regression, have smaller MAE and MSE. Thus, they were used in constructing the strategy.

After comparing the two models used in the strategy, we decided to use kernel regression as our prediction model because it performed better under our assumptions and strategy.



Kernel Model

Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	20.806%
Sharp Ratio	3.33
Max Draw Down	-1.556%



Robust Regression

Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	3.487%
Sharp Ratio	0.316
Max Draw Down	-2.787%

Figure 13 Portfolio performance comparison by using different prediction model

4. Long-short Strategy

4.1 Criteria Used in Construction of The Trading Strategy

4.1.1 Stock Selection

We used MAE and MSE results got from the validation set to weighted summarize a score to represent the accuracy of the model. The lower the score, the more precise it performs on predicting one stock.

Then we use the standard deviation of the predicted result together with the score of the MAE and MSE, to select the stocks with more accurate predicted result as well as a higher volatility in return rate.

4.1.2 Trading Signal

As for the trading signal, we mainly used thresholds generated from predicted mean and standard deviation. Since the stock return obeys the Geometric Brownian motion, according to the formula, the only thing to be considered about is the random variable, epsilon, which obeys the standard normal distribution.

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \epsilon \sqrt{\Delta t}$$

where:

$S =$ the stock price

$\Delta S =$ the change in stock price

$\mu =$ the expected return

$\sigma =$ the standard deviation of returns

$\epsilon =$ the random variable

$\Delta t =$ the elapsed time period

Figure 14 Geometric Brownian motion formula

The epsilon has over 68% probability to be larger than -1 and smaller than 1. It means that there are 68% that the stock return rate will be within one sigma from the mean return rate in a day. Hence, when it is larger than 1 or smaller than -1, we deemed that the return rate had exceeded the normal range, and that was our trading opportunity. Each selected stock has its threshold. We used the threshold of each stock to decide the stop loss and take profit points.

4.1.3 Market Trend Decision

The market trend was used to decide the adjustment of the parameters of the strategy. We used the VIX, IV, IV gap, yield, and crossing to decide market trend. The indicators might offer signals for several times within a month, we thought it will be too frequent to adjust the parameters of the strategy, leading to a worse performance. Thus, we adjusted our parameters only after the current parameters had been used for over 30 trading days.

4.2 Strategy Construction

4.2.1 Volatility Reduction

We used short position to hedge the long position. We could not only reduce the volatility of the portfolio but also minimize the max draw down. Because the long and short position are naturally hedging positions.

4.2.2 Stop Loss & Take Profit Points

Stop loss and take profit points were used to decide when to wipe out the position. They were calculated according to previous day's close and predicted return rate. The stop loss point might be modified according to open position price.

4.2.3 Trading Implementation

Mostly we trade at the opening price. But if the opening is too close to the take profit points, we will trade at previous day's price, or we do not trade at all.

We prefer intraday trade. If we cannot wipe out the position within a day, we will wipe it out in the coming days when either of the points are reached.

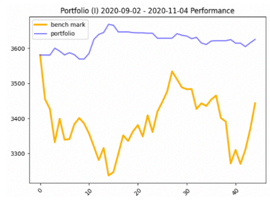
4.2.4 Transaction Costs

We considered explicit costs here. Since we need more market data, like trade data, minute scale data, to consider the implicit costs, we leave it the next development stage.

4.3 Back-testing

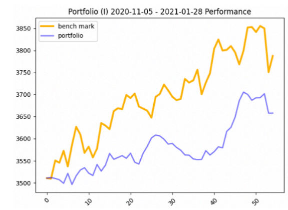
4.3.1 Different Time Horizon

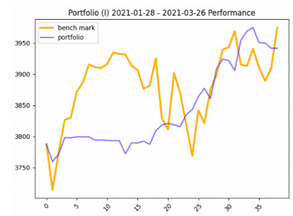
We implemented the strategy on different time zone, including the bullish and bearish. We found out that the strategy was stable, not worse than the benchmark if the benchmark dropped rapidly, but also no better than the benchmark if it rose sharply.



Start Date: 2020-09-02	End Date: 2021-11-04
Trading days	44
Annual Return	7.356%
Sharp Ratio	1.26
Max Draw Down	-1.748

Start Date: 2020-11-04	End Date: 2021-
Trading days	54
Annual Return	20.703%
Sharp Ratio	2.82
Max Draw Down	-1.564%



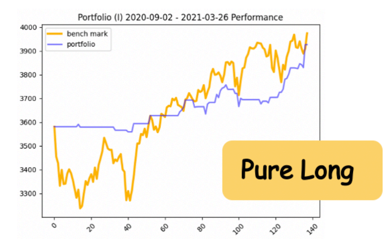


Start Date: 2020-01-28	End Date: 2021-03-26
Trading days	38
Annual Return	29.631%
Sharp Ratio	4.43
Max Draw Down	0.844%

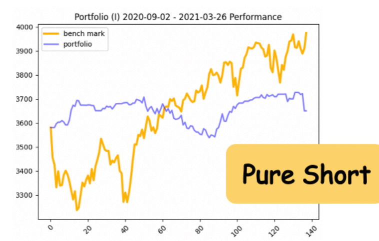
Figure 15 Portfolio performance in different period with various market conditions

4.3.2 Different Long Short Position Rate

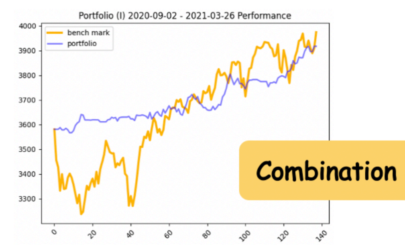
We tried pure long and pure short strategies, also tried strategies with different proportion. We found out that in bullish market, the more the long position would lead to a better performance, and vice versa. But pure long and pure short was not a good idea here. Because the combination strategy shows the best performance with higher sharp ratio and lower max draw down.



Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	17.616%
Sharp Ratio	2.64
Max Draw Down	-2.465%



Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	3.573%
Sharp Ratio	0.347
Max Draw Down	-4.671%



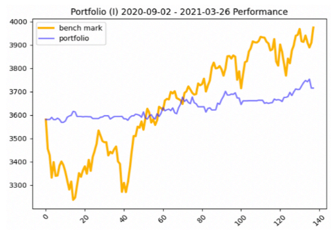
Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	17.147%
Sharp Ratio	2.85
Max Draw Down	-1.736%

Figure 16 Portfolio performance with different long short position rates

4.3.3 Different Aggressiveness

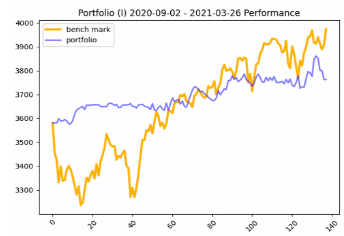
We added in aggressiveness as a parameter to adjust the take profit points. The larger, the more aggressive the strategy is, vice versa.

Neutral



Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	6.885%
Sharp Ratio	1.09
Max Draw Down	-1.776%

Aggressive



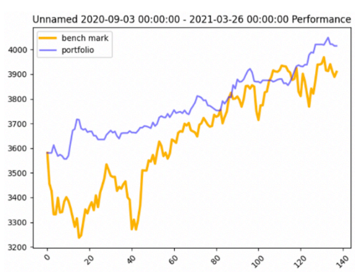
Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	9.297%
Sharp Ratio	1.13
Max Draw Down	-2.609%

Figure 17 Portfolio Performance with different aggressiveness

Usually, if the market is bullish market, the more aggressiveness is put on the long position and less on the short position. If the market is bearish, opposite ratio should be used. If the market is neither bull nor bear, then the aggressiveness should be equal in both sides. The number should be decided according to the volatility related indicators like VIX.

4.4 Dynamic Usage

If we dynamically adjusted the parameters of the portfolio, then the return of the portfolio might be better.



Start Date: 2020-09-02	End Date: 2021-03-26
Trading days	137
Annual Return	21.87%
Sharp Ratio	1.74
Max Draw Down	-2.225%

Figure 18 Dynamic investment with adjustment of parameters

Also, the modification of parameters should be taken only after the strategy had been launched for over 30 trading days owing to the advantage of the strategy occurred in longer term holding period. Moreover, each adjustment should be based on the current market situation.

4.5 Robo-advisor & Smart Investment

All in all, our strategy provides a stable profit within a long term. Although it cannot always beat the market if the market is in extreme bullish market, it performs better than the market under bearish market. Thanks to the natural hedge of the strategy, the max draw down of the portfolio value will not be too large.

What makes our fund a smart investment is that we offer suggested portfolio parameters under different market conditions.

5. Risk Management

There are various explicit and implicit risks faced by traders in the process of trading, including market volatility. Through efficiently identifying, analyzing, quantifying, and controlling these potential risks, traders can proactively mitigate and manage potential losses in an investment. Given every investment incorporates at least some degree of risk and return of an investment is inseparable from potential risks, our model attempts to manage these risks by quantifying the maximum loss possible as well as understand the return compared to the risk.

5.1 Maximum Drawdown (MDD)

A maximum drawdown (MDD) is the maximum observed loss from a peak to a trough of a portfolio during any investment before a new peak is achieved. As an indicator of the magnitude of the downside risk over a specified period, MDD not only measures the historical volatility but also provides means for predicting future price movements and future risks. The lower the maximum drawdown the better given less potential loss. Our portfolio had a maximum drawdown of -1.56% from the peak date of Dec 15, 2020, to the trough date of De 31, 2020.

$$MDD = \frac{Trough\ Value - Peak\ Value}{Peak\ Value}$$

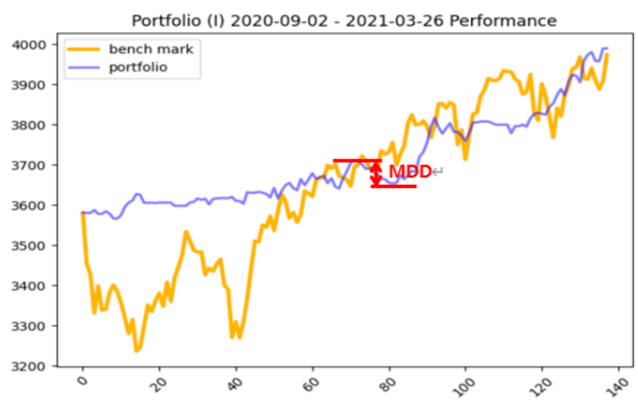


Figure 19 Max Draw Down of the strategy performance

5.2 Risk-adjusted Return Statistics

Our model not only quantifies the maximum loss in investment, but also considers risk-adjusted return statistics.

5.2.1 Sharpe Ratio

Sharpe ratio is a measure of risk–adjusted performance indicating the level of excess return per unit of risk, whereby measuring volatility with the standard deviation. By subtracting the risk-free rate from the mean return, investors can isolate the profits associated with risk-taking activities. A Sharpe ratio of 1.0 or greater implies that for every unit of risk, investors assume to achieve an equal amount of return. Our portfolio has a Sharpe ratio of 3.33.

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

where:

R_p = return of portfolio

R_f = risk-free rate

σ_p = standard deviation of the portfolio's excess return

Figure 20 Sharpe Ratio formula

5.2.2 Calmar Ratio

The Calmar ratio is a risk-adjusted performance measure that considers the drawdown to assess the performance. It is a function of the annualized rate of return over the absolute value of the maximum drawdown, usually over a period of three years. Generally, both the annualized return and the maximum drawdown are based on the last 36 months of monthly returns. The higher the Calmar ratio, the better it performed on a risk-adjusted basis during the given time frame. Although it is difficult to define what a good Calmar ratio is, it is often believed that a Calmar ratio of 3.0 to 5.0 is deemed good. Our portfolio has a Calmar ratio of 13.3 due to very low MDD.

$$\text{Calmar Ratio} = \frac{Annualized\ Return\ (R_p)}{|MDD_p|}$$

where:

R_p = Annualized return of portfolio

MDD_p = Maximum drawdown of portfolio

Figure 21 Calmar Ratio formula

6. Potential Improvements

6.1 Access to Better Quality Data Source

Our strategy currently uses daily data including open, close, high, and low price. To further develop our strategy, we plan to include more data, such as near real-time intraday data to improve the strategy.

Having intraday source of data in minutes, seconds scale, or even real-time data including the tick data, the trade data, and the order book visibility would enhance the accuracy of our model. We could use orderbook signals to predict trends more precisely, enhance the entry points of our positions dynamically throughout the day, and further develop intraday trading strategies.

6.2 Developing Intraday Trading Strategies Based on New Benchmark

With more precise intraday data mentioned above, we can have more frequent intraday trading strategies instead of one to two transactions of each stock. Our model can be adjusted to test out various benchmarks such as VWAP, open price, arrival price (IS) or close price with different intraday algo strategies to see which algo provides the best performance versus our benchmark.

6.3 Hedging with Different Financial Products

Currently our model tries to manage risks through various means such as taking long/short positions, implementing stop loss limits and equally weighting portfolio on different positions. Risk could be further managed by using other financial products such as futures or options to delta hedge our positions.

6.4 Rolling Basis Training, Validating, and Testing

Currently our model trains, validates, and back tests based on below fixed schedule. Our model could be further improved by attempting to train, validate and back test our strategy on a rolling basis as time progresses, instead of using fixed period. Having a model running on rolling basis would not only enhance the accuracy and preciseness but also ensure that our strategy works in various market conditions and different scenarios.

Figure 22 Rolling basis model training, validating, and testing

References:

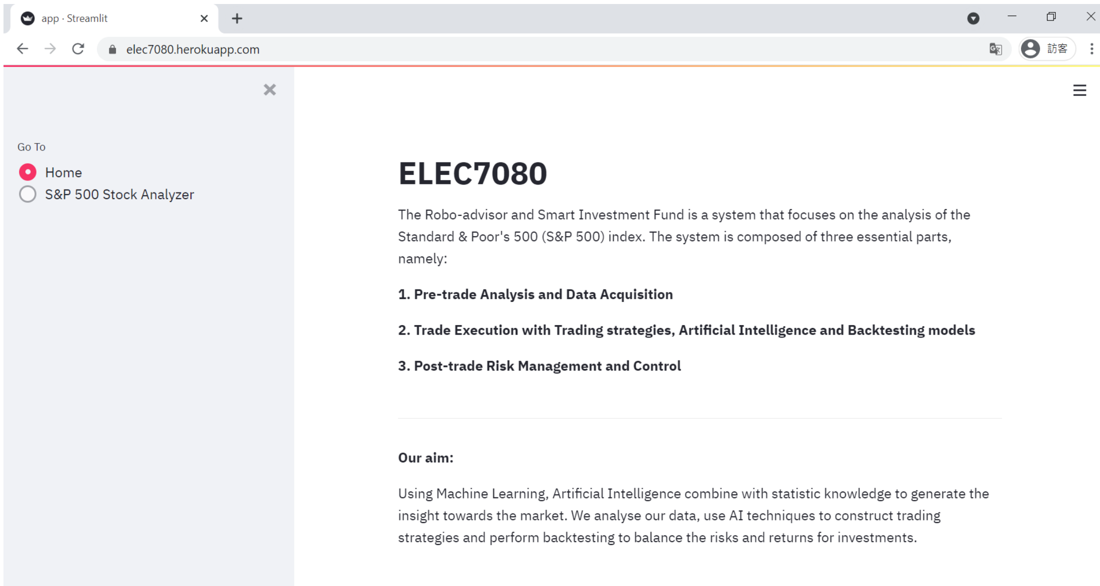
[1] Art Paspanthong, Divya Saini, Joe Taglic, Raghav Tibrewala, Will Vithayapalert. Trend Following: A Machine Learning Approach[D].California.Stanford University, 2019.

[2] Angelos Filos.Reinforcement Learning for Portfolio Management[D].London.Imperial College London, 2018.

[3] Haoran Wang, Xun Yu Zhou.Continuous-Time Mean–Variance Portfolio Selection: A Reinforcement Learning Framework[D].Columbia University, 2019.

Appendix:

Home page:



S&P 500 Stock Analyzer page: