The University of Hong Kong

Department of Computer Science
FITE4801 Final Year Project (2023-24)

FYP23003: Discovering Winning Strategies in Algorithmic Trading through a Comprehensive Approach

Detailed Project Plan

Lee Sze Choi (3035779571)

Li Wang Hei (3035782279)

Tsang Yan Chak (3035780790)

Under the supervision of Prof S. M. Yiu

Contents

1. Project Background

The emergence of algorithmic trading dates to the 1970s, where basic if-then-else algorithms are used for trade executions. As the Internet become more advanced, financial markets started to widely adopt digitalized execution and electronic communication networks (ECNs) in the 1990s and 2000s (McGowan, 2010). In recent decades, the use of machine learning and artificial intelligence in high frequency trading has been popularized.

Complex mathematical models and technical indicators are used to create signals for trade execution, which is then conducted by algorithms in the financial market. Algorithmic trading had been gaining popularity in the past decades, and different strategies were created and proven to be profitable. However, given the competitive nature of the algorithmic trading industry and the theories developed by the *Efficient Market Hypothesis*, previously profitable strategies will be exploited and returns will diminish.

In this project, we aim to create a new winning strategy that can cater for different market situations and outperform previously known strategies. This report will cover literature in different areas of algorithmic trading and multiple strategies. It will then outline the objectives and methodology of our project, and finally conclude with the targeted timeline of the project.

2. Literature Review

Nuti et al. (2011) identified different types of algorithmic trading systems, which are classified into execution trading systems, market-making systems and proprietary trading systems. This has laid the foundation of clear classifications between different algo trading systems. Nuti et al. (2011) then identified the similar stages of flow in different trading systems, which are the pre-trade analysis, trading signals and trade execution stages. However, they did not explore and compare different strategies in their research.

While a lot of literatures review and compare strategies that focus on execution trading and high-frequency trading (HFT), fewer literature focuses on comparing different arbitrage and non-HFT proprietary algorithmic trading systems. This phenomenon is caused by the current trading volume being dominated by HFT trading algorithms and execution trading algorithms. NASDAQ (n.d.) estimated that 50% of trading volume is driven by algorithmic high frequency trading systems in the United States stock market. Execution trading systems are also a focus of algorithmic trading researches. Fraenkle and Rachev (2009) have researched different market microstructures, benchmarks, patterns in trading volume analysis used in execution algorithmic trading systems. However, there is no comparison between the performance or trading cost used in different trading systems.

Researchers also focus on creating or analyzing specific algorithmic trading strategies. For example, Chu et al. (2020) have researched high-frequency momentum trading strategy, combining exponential moving averages with different durations on cryptocurrencies. Lv et al. (2019) have researched different machine learning and deep learning algorithms, using 44 technical indicators as input to predict stock prices. The research used different metrics including the annualized Rate of Return, Sharpe Ratio, Win Rate and Maximum Drawdown (Lv et al., 2019), setting a reference for using similar metrics in this research. The research found that machine learning algorithms produce higher returns and Sharpe ratios comparing to the index (Lv et al., 2019). However, it focused purely on the US stock market and the Chinese A-shares market; it also did not take transaction costs into consideration.

3. Project Objectives

The first objective of this project is to evaluate different algorithmic trading strategies and examine their performance on predicting price actions across the equity and cryptocurrency market. This project aims to research multiple strategies across different markets to investigate whether the strategies work across markets while taking transactional costs into consideration, and the reasons behind different results, which very few research have done so. Then, the second objective is to plan to combine different winning strategies to make a strategies that will work in most environment and diversified in different markets if possible.

During the initial phase, the focus will be on optimizing the performance and efficiency of the best performing strategies. After analyzing the backtesting results of existing strategies, this project will attempt to create a profitable strategy by combining the best strategies. This project will predominantly utilize trading, on-chain and alternative data to develop trading strategies.

This project also aims to compare and contrast the performance of different algorithms in the US, HK and cryptocurrency market. US market is considered as the mature equity market while HK market is a less developed market in terms of algorithmic trading. On the other hand, cryptocurrency is a new market for algorithmic traders.

Apart from that, this project will emphasize on evaluating the strategies' performance during black-swan events and over a longer time horizon to provide a holistic view on the effectiveness of the reviewed strategies, while taking transactional costs into considerations. This project will provide a comprehensive report on trading algorithms by examining both the existing and newly developed strategies in markets that are in different development stages.

4. Project Methodology

This project has several key aspects to focus on, including data collection, backtesting, modelling, metrics evaluation, and front-end development.

4.1. Data Collection and Backtesting

Regarding data collection, the project requires past stock prices and returns as the first step to building predictive models. The two major sources of financial data are from (1) the QuantConnect backtest engine, and (2) the Yahoo Finance API. This project will then perform data cleansing, preprocessing and aggregation using Python and the Pandas library. This project will perform backtesting using the platform provided by QuantConnect.

4.2. Modelling

In terms of modelling, the project will make use of the following models to predict the security prices and their buy/sell signal:

1. Traditional approaches, such as momentum trading and arbitrage;
2. Event-driven strategies, such as earnings;
3. Classical time-series analyses, such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditionally Heteroskedastic (GARCH) models;
4. Machine learning approaches, such as Random Forest, Support Vector Machines (SVM) and Logistic Regression;
5. Deep learning approaches, such as Convolutional Neural Networks (CNN) and Natural Language Processing (NLP) techniques.
6. Other models researched or created, such as statistical patterns or dark volume related strategies.

4.3. Metrics Evaluation

The project will focus on the following metrics to evaluate the strategies:

1. *The Annualized Rate of Return (ARR)*

   The annualized rate of return is a common evaluation for fund performance, which is derived from net returns. The net return function is as follows, assuming no distribution and reinvestments (Cuthbertson et al., 2010).

   $$1 + R_{net}^{t} = \frac{NAV_t}{NAV_{t-1}}$$

   The net return (R_{net}^{t}) is the overall return from t-1 to t. As different strategies may have different evaluation periods due to data availability, the net return is then adjusted to annualized rate of return.

   $$R_{annualized} = (1 + R_{net}^{t})^{\frac{1}{n}} - 1$$

   The annualized rate of return ($R_{annualized}$) will be adjusted according to the number of years (n), making the evaluation metric more fair.

2. *Win Rate*

   The win rate is the percentage of trades that have produced a positive return, a metrics to analysis the competence of the trading strategies disregarding the return of the trades. The win rate function is as follows.

   $$\text{Win rate} = \frac{\text{Number of trades with positive return}}{\text{Total number of trades}}$$

3. *Sharpe Ratio*

   Sharpe Ratio is widely used for evaluating fund performance, particularly by finance practitioners (Cuthbertson et al., 2010). It measures the reward-to-risk ratio, where risk is defined by the variance of return of the portfolio (Cuthbertson et al., 2010). The formula of Sharpe Ratio is as follows.

$$\text{Sharpe Ratio} = \frac{R_{net} - R_f}{\sigma_R}$$

The risk-free rate (R_f) will the rate of return of risk-free assets, which will be estimated via U.S. Treasury Bond, adjusted for credit default risks.

4. *Maximum Drawdown (MDD)*

   Maximum Drawdown is a common metrics used to evaluate hedge fund, which captures the maximum peak-to-trough return over the entire duration of the fund (Van Hemert et al., 2020). This project also employs the same metric to measure the largest drawdown from the highest historical return over the entire evaluation period.

5. *Alpha*

   Alpha is another common metric used to find the excess return above the benchmark (Cuthbertson et al., 2010), measuring the ability of the trading strategies to generate unsystematic returns. Systematic returns could be generated through investing into the market portfolio or benchmark portfolio (m), which does not require developing algorithmic trading strategies. However, unsystematic return, measured by alpha, could not be generated from market portfolio or benchmark portfolio. Therefore, alpha, which is the intercept term in the regression formula shown below, reflects the value and skill of the algorithmic trading strategies (Cuthbertson et al., 2010).

$$\alpha = R_{net} - [R_f + \beta(R_m - R_f)]$$

   In the formula above, R_f is the risk-free rate, R_m is the market return or benchmark return and β is the portfolio beta. The portfolio beta is the expected movement of the portfolio relative to the movement in the market or benchmark portfolio. Portfolio beta is calculated as follows (Cuthbertson et al., 2010).

$$\beta = \frac{Cov(R_{net}, R_m)}{Var(R_{net})}$$

In this project, the benchmark will be chosen according to the market under investigation. For the stock market, the S&P 500 will be chosen as the benchmark. For the cryptocurrency market, the S&P Cryptocurrency LargeCap Index will be chosen as the benchmark. For mixed markets, the portfolio will be tested against both benchmark for references.

Overall, this project aims to maximize ARR, Win Rate, Sharpe Ratio and Alpha, while minimizing MDD.

4.4. Frontend Development

Finally, the results of the project will be displayed in a web dashboard. The project is tentatively using Express.JS as the framework for displaying the returns and profits of various strategies.

5. Feasibility Research

In the past month, our members have studied the feasibility of 15 strategies and backtested them, including (i) momentum trading using moving averages, (ii) GARCH model, and (iii) machine learning strategies.

5.1. Momentum Trading

The simple moving average trading strategy utilized the intersection of stock price with the moving average (MA) to generate trading signals. This strategy expected that stock price momentum continues after the price crosses the MA. It initiates a long trade when the price crosses the MA upwards by more than a certain percentage and initiates a short trade when the price crosses the MA downwards by more than a certain percentage. In addition, a trailing stoploss is set to limit our loss and a target price is set to take profit. After backtesting against the period from January 2019 to January 2023, this algorithm obtained a Sharpe ratio of 2.44 and an annualized return of 91.25%.

5.2. GARCH Model

Generalized Autoregressive Conditionally Heteroskedastic (GARCH) model is one of the commonly used statistical methods to model financial time series, accounting for the volatility of the underlying series. This strategy predicts the next log-return of the series, given the current log-differenced series of equity prices. During the prediction stage, the price is predicted by exponentiating the return series, multiplied by the last observed price. Compared to other models, a GARCH process can capture the underlying fluctuations due to the phenomenon of volatility clustering (Bai et. al., 2003), which gives a more accurate point prediction. Overall, the best-performing model has obtained a Sharpe ratio of 1.15, an annualized return of 39% and a CAPM alpha of 0.18.

5.3. Machine Learning Algorithms

Machine Learning algorithms have been prevailing in its use in algo trading (Hansen, 2020). The feasibility study has covered basic machine learning models to study its base performance in algo trading. The algo trading model used is based on 6 different technical indicators, namely last 1 Day price change, 50 days simple moving averages (SMA), 100 days SMA, 200 days SMA, RSI and Moving Average Convergence/ Divergence (MACD). These technical indicators are then trained to predict the forward 5 days return. Different machine learning models are being trained and tested, including k-nearest neighbor (KNN), support vector machine (SVM)[1], decision tree (DT), random forest (RF), logistic regression (LR) and gradient boosting (GB). The model was trained and tested on different stocks and indices, the following post-transactional cost result shown in *Table 1* is based on SPY (an exchange traded fund tracking S&P 500 index)[2].

Metric	**S&P500**	**KNN**	**SVM**	**DT**	**RF**	**LR**	**GB**
Annualized Rate of Return	11.07%	0.70%	1.26%	4.16%	3.39%	1.64%	7.87%
Win Rate	N/A	71%	100%	64%	82%	55%	61%
Sharpe Ratio	0.563	0.248	0.406	0.482	0.664	0.185	0.532
Maximum Drawdown	33.5%	32.6%	2.6%	12.0%	7.9%	129.5%	22.3%

Table 1. Backtesting results from various machine learning algorithms trading on SPY from December 2015 to January 2023

Although most of the machine learning models are having a lower return and Sharpe ratio then the index (S&P 500), this laid a foundation of machine learning models able to generate trading signals. This is a proof-of-concept of basic machine learning's ability to predict stock prices movement based on limited amount of information. Moreover, most machine learning models have a much lower maximum drawdown than the index. In the future, this research could be extended to deep learning and other machine learning models, the input data and technical indicator list could also be extended.

1. Note that the win rate for SVM in Table 1 is 100 %. This is due to the fact that the algorithm had only made 2 trades.

2. Note that the win rate for S&P 500 is not available as it is the benchmark index.

6. Challenges of the Project

There are several challenges the project may encounter during its development, such as the availability of data, the complexity of models and the difficulty of developing own models under the time constraint.

6.1. Data Availability

There are limited data sources for algorithmic trading for the Hong Kong stock market due to a low participation of retail trading in algorithmic trading in Hong Kong. Moreover, there are also limited availability of tick/ trade level data with conditional codes of the trade, where some of the trading strategies require very granular data (i.e. tick-level data) and specific conditional codes.

However, this project will source from different data sources, if possible, to increase the availability of data sources. The major data source for US stock data and Crypto data will be sourced from QuantConnect's Researcher plan. HKEX and WRDS database will also be used, if available, for Hong Kong stock data and options data. However, if such database is used, a backtesting engine will have to be developed on our own, adding complexity to the current project.

6.2. Model Complexity

Some deep learning models will require a long time and huge computational power to train. To optimize hyperparameters and strategies, these models will have to be trained and tested repetitively.

The HKU CS GPU farm or QuantConnect's research node may be used for increasing the speed of training the model and optimizing the hyperparameters. Moreover, this project will also reference to researches that have previously optimized the models, in order to shorten the time needed to optimize the hyperparameters.

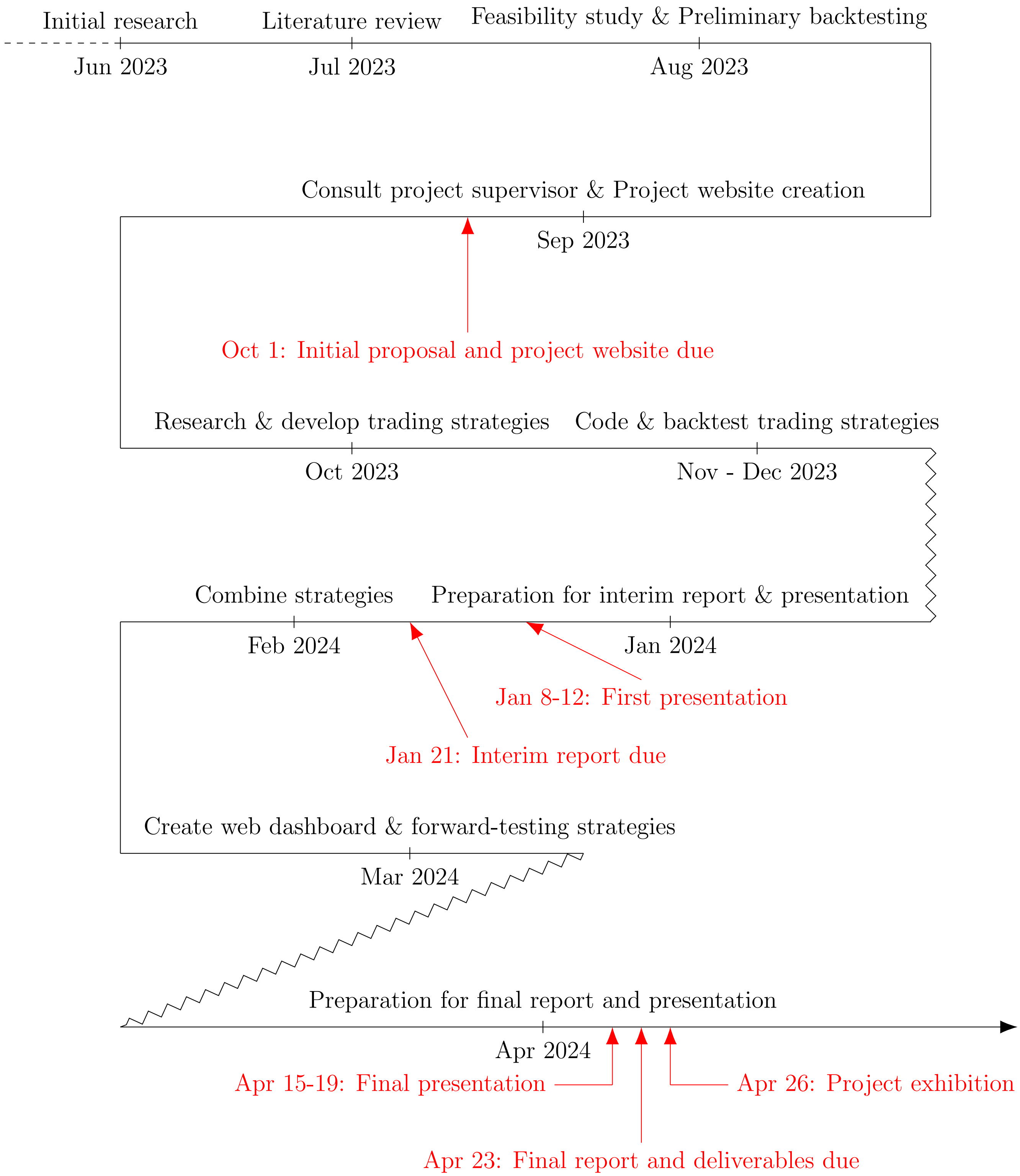
6.3. Difficulty in developing own models

Developing trading strategies that are unique and unresearched are challenging. Most of the trading strategies could not produce high return, Sharpe ratio and positive alpha. This project will focus on models through statistical patterns identified and patterns identified through analyzing different microstructures in different markets.

With different challenges and time constraints of the research, this project may reduce the number of markets under research and the number of trading strategies to be researched.

7. Project Schedule and Milestones

As of September 2023, the tentative schedule for the project is shown below:



There are several deliverables throughout the project, which are shown in *Table 2*:

Item	**Deliverable(s)**	**Month**
1	Feasibility Report	Aug 2023
2	Project proposal for supervisor	Sep 2023
3	Detailed project plan	Sep 2023
4	Project webpage	Sep 2023
5	Backtesting results for various trading strategies	Sep - Dec 2023
6	Preliminary strategy implementation	Jan 2024
7	Detailed Interm Report	Jan 2024
8	First presentation slides	Jan 2024
9	Backtesting results for combined strategies	Feb 2024
10	Forward testing results for combined strategies	Mar 2024
11	Web Dashboard	Feb - Apr 2024
12	Final presentation slides	Apr 2024
13	Final tested strategy implementation	Apr 2024
14	Final report	Apr 2024

Table 2. Project Deliverables

References

Bai, X., Russell, J. R., & Tiao, G. C. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. Journal of econometrics, 114(2), 349-360.

Chu, J., Chan, S.,& Zhang, Y. (2020). High frequency momentum trading with cryptocurrencies. Research in International Business and Finance, 52, 101176. https://doi.org/10.1016/j.ribaf.2019.101176

Cuthbertson, K., Nitzsche, D.,& O'Sullivan, N. (2010). Mutual fund performance: Measurement and Evidence¡sup¿1¡/sup¿. Financial Markets, Institutions& Instruments, 19(2), 95-187. https://doi.org/10.1111/j.1468-0416.2010.00156.x

Fraenkle, J.,& Rachev, S. T. (2009). Review: algorithmic trading. Investment Management and Financial Innovations, 6(1), 7-20. https://www.researchgate.net/publication/283166465_Review_Algorithmic_trading

Hansen, K. B. (2020). The virtue of simplicity: On machine learning models in algorithmic trading. Big Data& Society, 7(1), 205395172092655. https://doi.org/10.1177/2053951720926558

Lv, D., Yuan, S., Li, M.,& Xiang, Y. (2019). An empirical study of machine learning algorithms for stock daily trading strategy. Mathematical Problems in Engineering, 2019, 1-30. https://doi.org/10.1155/2019/7816154

Michael J. McGowan, The Rise of Computerized High Frequency Trading: Use and Controversy, 9 DUKE L. TECH. REV., 2010, at 1, 4-7.

NASDAQ. (n.d.). High frequency trading (HFT). https://www.nasdaq.com/glossary/h/high-frequency-trading

Nuti, G., Mirghaemi, M., Treleaven, P.,& Yingsaeree, C. (2011). Algorithmic trading. Computer, 44(11), 61-69. https://doi.org/10.1109/mc.2011.31

Quantified Strategies. (2023, August 26). What percentage of trading is algorithmic? (Algo trading market statistics: Growth, trends, and forecasts). Trading Strategies - Quantified Strategies. https://www.quantifiedstrategies.com/what-percentage-of-trading-is-algorithmic/

Van Hemert, O., Ganz, M., Harvey, C. R., Rattray, S., Sanchez Martin, E.,& Yawitch, D. (2020). Drawdowns. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3583864