MATH 4999 - Independent Capstone Project
Statistical and Machine Learning Research in Financial Market

Author: Leung Pak Hei, Marco
Advisor: Prof. Kani Chen

Department of Mathematics, The Hong Kong University of Science and Technology

December 23, 2022

1 Introduction

The stock market is one of the major areas that financial specialists are interested in, hence stock market price slant expectation is continuously a hot subject for analysts from both financial and technical spaces. Nowadays, Automatic Trading is being adopted by both hedge funds and retail investors. Even traditional players like Investment banks and portfolio managers are also switching from manual trading to Quantitative trading. According to the Yahoo Finance report, it is estimated that around 70% to 80% of shares traded on U.S. stock exchanges are driven by the automatic trading. [7] With the more mature development of artificial intelligence, it is possible to combine the statistical and machine learning model with the automatic trading bot to generate huge profit. Also, Statistical research and machine learning requires a lot of data. Due to the openness of the financial market, historical data and real-time data can be easily assessed through data providers or stock brokers to provide sufficient data for our research. In the first part of this research, our objective is to do statistical research on the financial market including the stock price momentum or some macroeconomics indicators to find meaningful nature of the stock Market, Then, trading strategies will be built based on several machine learning models (CNN/DSTM/Deep Learning, etc) and back-test it to see the performance of the strategies. Finally we would compare the performance of different and discuss its pros and cons and the limitations and improvements.

2 Literature Review

2.1 Automatic Trading

Algorithmic trading permits traders to set up rules for both trade entries and exits that, once modified, can be consequently executed through a computer. In reality, different stages report that 70% to 80% or more of stocks in the global market to come from programmed algo-trading bot. One of the greatest attractions of algorithmic trading is that it can take emotion out of trading since trades are consequently controlled by the objective trading rules.[7]

2.2 Financial Market

2.2.1 Stock Market

The stock market alludes to trades in which shares of publicly held companies are traded. Monetary activities are conducted through formal trades and by means of over-the-counter (OTC) marketplaces that work beneath a characterized set of regulations. Nasdaq and New York Exchange are few of the American exhchanges that provide stock trading.

2.2.2 Derivative Market

The derivatives market refers to the financial market for financial instruments such as Forward contracts, Swap contracts, , Warrants, futures contracts or options that are based on the values of their basic assets. Chicago Mercantile Exchange (CME), International Securities Exchange (ISE), the Intercontinental Exchange (ICE), Hong Kong Futures Automated Trading System (HKATS) and the LIFFE exchange are few of the examples that provide derivatives products.

2.2.3 Forex Market

Forex (FX) is a portmanteau of foreign currency and exchange.Due to the demand from trade, commerce, and finance, forex markets tend to be the asset with largest liquidity and trading volume. Hedgers, institutional traders use strategies related to Forex to hedge against crisis such as interest rate risk, geopolitical events, and to diversify portfolios.

2.2.4 Crypto Market

A cryptocurrency is a shape of digital asset based on an organization that's dispersed over a huge number of validators and miners. This decentralized structure permits them to exist exterior the control of governments and central specialists.
The Crypto Market can be seen in both centralized platforms and decentralized platforms. Centralized exchange, such as Binance and FTX, is usually operated by a company with central authority in which they can have the full information on the users and the trading systems. On the other hand, decentralized exchange, such as Uniswap and PancakeSwap, is a operated on a decentralized network in which everyone can be the validators or decision makers to validate the transaction and vote on the policy change.

2.3 Economic Indicators

2.3.1 Gross Domestic Product

Gross domestic product (GDP) is the mathematical presentation of the production of goods and services in a country over a time-series. It follows the following three types of approaches:

Expenditure Approach: $GDP = Consumption(C) + InvestmentSpending(I) + GovernmentSpending(G) + NetExport(X - M)$

Income Approach: $GDP = NationalIncome + Salestax + Depreciation + NetForeignFactorIncome$

Value-added Approach: $GDP = Output - IntermediateConsumption$

2.3.2 Federal Interest Rate

In the US, the federal funds rate is the interest rate at which banks and credit unions lend reserve balances to other institutions overnight. Reserve balances are required to maintain the reserve requirements. The federal funds rate is an important benchmark in financial markets. Theoretically, it is highly correlated with the currency rate, inflation rate, real estate price and the stock market sentiment[13].

2.3.3 Consumer Price Index (CPI) / Inflation rate

The Buyer Cost File (CPI) measures the changes over time within the price level of buyer merchandise and administrations by and large obtained by households. The year-on-year rate of alter within the CPI is broadly utilized as an pointer of the expansion influencing customers.

2.3.4 Commodities Price

Most commodities are raw materials, fundamental assets, agrarian, or mining items, such as press mineral, sugar, or grains like rice, wheat and chemicals. Prevalent commodities incorporate crude oil, corn, and gold. The price of a commodity good is decided by either well-established physical commodities have effectively exchanged spot and subsidiary markets. The wide accessibility of commodities regularly leads to littler benefit edges and lessens the significance of factors other than price. But since the commodities are the raw materials of the industrial activities. If the commodities change rapidly, it can affect the financial market as a whole in a significant extent. [5]

2.4 Trading Strategies

2.4.1 Random Flipcoin Strategy

For Each day, it is traded either long and short with equal probabilities. The profit for each day is calculated by: $(close - open)/open * size$

2.4.2 Pairs Trading

Pairs trading is a non-directional, relative value speculation methodology that looks for to recognize 2 assets with comparative characteristics whose value securities are currently exchanging at a cost relationship that's out of their authentic exchanging run. For example, the airline companies within the same sector such as AAL, UAL and DAL should have similar stock price movement as they are affected by similar macroeconomics. Theoretically, the only difference is mostly come from the "systemic risk" of the stock from the management and companies investment. Therefore, we expect there will be a "convergence" and "divergence" a stock which creates arbitrage opportunities [15].

2.4.3 Asset Allocation strategy

Asset allocation could be a methodology to balance risk and returns by contributing totally different asset types. Historical price trends of diverse asset classes appear moo or negative relationship among these asset classes. Hence diversification over asset classes can significantly decrease risk and create potential predominant returns within the long term. In general, there are three types of asset allocation, with Strategic Asset Allocation, Tactical Asset Allocation and Dynamic Asset Allocation.

2.4.4 Momentum strategy

Momentum Trading Technique is a technique in which financial specialists long a position that's rising and sells them when they see to have crested. The objective is to work with volatility by finding longing opportunities in short-term up-trends and after that sell when the securities begin to lose momentum. At that point, the speculator takes profit and looks for another short-term uptrend, or buying opportunity, and rehashes the strategies. Compared to other trading strategies, Momentum strategy is prediction-oriented and usually depending on some traditional indicators like price, volume or some technical indicators like RSI, MACD.

2.4.5 News/Sentiment strategy

Benoit Mandelbrot proposed the efficient markets hypothesis (EMH) and claims that markets are efficient, leaving no opportunity to make excessive returns by investing as the information and the price of the asses are already priced in. In addition, prices in financial markets are impartial and contain all the shrewdness or future estimates from financial specialists [4]. Assuming everyone is subjective and rational, no one can beat the market. However, with the rise of the volume trading from the retail investors, we can always see there is a momentum driven from the "sentiment" and more and more irrational behaviors. Also, in reality, the information is sometimes, if not always, not perfectly price and closed. There is always a time-lagging for the news to be fully fairly priced. For example, the Wall Street Bet forum formed the short squeeze in the MeMe stock. The pessimistic news and speech by the media drive the selloffs of the stock market in 2022. These fearfulness or greediness of the stock market can create a temporarily misprice opportunities for the speculators to take advantages.

2.5 Statistical Knowledge

2.5.1 Heteroscedasticity

In insights, heteroskedasticity happens when the standard deviations of a anticipated variable, observed over diverse values of an autonomous variable or as related to earlier time periods, are non-constant. With heteroskedasticity, the tell-tale sign upon visual assessment of the residual errors is that they will tend to fan out over time.
Mathematical Speaking, Heteroscedasticity occurs when: $Var(A \mid B) \neq const$

2.5.2 Autocorrelation

Autocorrelation is the correlation with a delayed copy of itself in a time series.The analysis of auto-correlation is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise, or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies. Auto-correlation coefficient can be calculated as:
$\hat{r_k} = \frac{\sum_{t=k+1}^{T}(y_t - \hat{y})(y_{t-k} - \hat{y})}{\sum_{t=1}^{T}(y_t - \hat{y})^2}$

2.5.3 Cointegration

Cointegration is a statistical strategy utilized to test the relationship between two or more non-stationary time series within a indicated time period. The strategy makes a difference in recognizing long-run parameters or balance for two or more sets of factors. It makes a difference in deciding the scenarios wherein two or more stationary time series are cointegrated in such a way that they cannot leave much from the balance within the long-run.

2.5.4 ARIMA

An autoregressive coordinates moving normal demonstrate (ARIMA) could be a form of regression analysis that gages the quality of one subordinate variable relative to other changing factors. The model's objective is to anticipate the future by looking at the differences between values within the arrangement rather than through real values.
An ARIMA model can be understood by outlining each of its components as follow:

1. Autoregression (AR): a changing variable that relapses on its lagged, or earlier, values.
2. Integrated (I): speaks to the differences of raw perceptions to permit for the time arrangement to ended up stationary.
3. Moving Average (MA): consolidates the reliance between an perception and a residual error from a moving average model connected to lagged perceptions.

2.5.5 Markov decision random process

Markov Decision Process is a common mathematical models that can be used to analyses the problems where the outcomes are random and controllable. Mackov Decision prcoess can exist in different forms. But they should mostly contains four parts: state space, action space, intermediate reward, policy funciton and optimization.

1. State Space: A set of states , denoted as S
2. Action State: A set of actions, denoted as A
3. Intermediate Reward $R_a\ (s, s')$: The reward generated from state s to state s' by action a as a short-term utility for the action.
4. Discount factor γ : A discounting effect/ inflationary effect for the future state space
5. Assumption: s_{t+1} and r_{t+1} rely solely on s_t and a_t.

2.5.6 Markowitz model

Markowitz model may be a portfolio optimization model which helps within the determination of the foremost effective portfolio by analyzing distinctive securities by choosing securities that don't change totally the same. It is based on two assumptions that the investors would prefer a portfolio with the highest expected mean (return) and lowest standard deviation (Risk). The model follows the equation: $R_P = R_{RF} + (R_M - R_{RF})\ \sigma_P\ /\ \sigma_M$

2.6 Machine Learning model

2.6.1 Convolutional neural network (CNN)

CNNs are regularized forms of multilayer perceptrons. Multilayer perceptrons ordinarily cruel completely associated systems, that's, each neuron in one layer is associated with all neurons within another layer. The "total network" of these systems makes them inclined to overfit information. Ordinary ways of regularization, or avoiding over-fitting, incorporate: penalizing parameters amid preparing or trimming the network. CNNs take a distinctive approach towards regularization: they take advantage of the various leveled design in information and collect designs of expanding complexity utilizing littler and less difficult designs decorated in their channels [9]. Subsequently, on a scale of network and complexity, CNNs are on the lower extraordinary.

2.6.2 Long short-term memory (LSTM)

As the the name of LSTM suggested, LSTM refers both "long-term memory" and "short-term memory". Unlike standard feed-forward neural systems, LSTM has input associations. Such a repetitive neural organize (RNN) can handle not as it were single information focuses but too whole information groupings. For real-world appllication, LSTM is pertinent to errands such as unsegmented, associated penmanship acknowledgment, discourse acknowledgment, machine interpretation, robot control, video recreations, and healthcare.
LSTM systems are well-suited to classifying, preparing and making forecasts based on time arrangement information, since there can be slacks of obscure term between imperative occasions in a time arrangement. LSTMs were created to bargain with the vanishing angle issue that can be experienced when preparing conventional RNNs. Relative heartlessness to crevice length is an advantage of LSTM over RNNs, covered up Markov models and other grouping learning strategies in various applications [1].
A common LSTM model consists of four components:

1. Cell
2. Input Gate
3. Output Gate
4. Forget Gate

2.6.3 K-means

K-means clustering is a unsupervised learning algorithm that performs the separation of information into non-overlapping clusters that offers similitude. The term "K" means the number of clusters required to form. The distance is calculated by the D between each point of the data to every centroid [2]. The commonest distance metrics used is the Euclidean distance: $D_{(p,q)} = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$

2.6.4 OPTICS

OPTICS is a density-based algorithm, which is comparative to the Density-based spatial clustering of applications with clamor (DBSCAN), but it can distinguish clusters in shifting density. OPTICS moreover considers focuses with more thickly stuffed clusters, so each point is doled out a center remove as the MinPts is the closest point. The Reachability Remove is characterized relative to another information point q(Let). The Reachability separate between a point p and q is the most extreme of the Center Remove of p and the Euclidean Remove between p and q [14]. If two focuses (p,q) are the closest neighbors, it is anticipated that they have a place to the same clusters.

3 Data Used

3.1 Target

1. Stocks under Russel Index 2000
2. Top 20 largest Crypto-currencies
3. Foreign Exchange Market
4. US T-Bond

3.2 Time Period

The financial in the past can perform very differently from nowadays. The assumptions of " the movement happens in the past will also happen in the future" may not hold especially in the technological stock and cryptocurrency. Therefore, the project will only be tested based from the historical data from 2015 to 2022. Strategies will be tested based on different sub-time interval.

3.3 Time Interval

For the simplicity of the research, we would conduct the research based on "day" interval based on candle chart data, consisting of "High", "Low", "Open", "Close".

3.4 Data provider

Yahoo Finance and Interactive Brokers

4 Programming development

4.1 Programming Language

In this project, R and STATA would be used for the statistical research and Python would be used for the machine learning modeling and backtesting.

4.2 External Programming platform and library

In this project, Tensorflow, Keras and PyTorch would be used for the machine learning modeling.

5 Statistical Research Results

5.1 Fundamental Data analysis of stock market

Figure 1: Price of SPY from 2007 to 2022

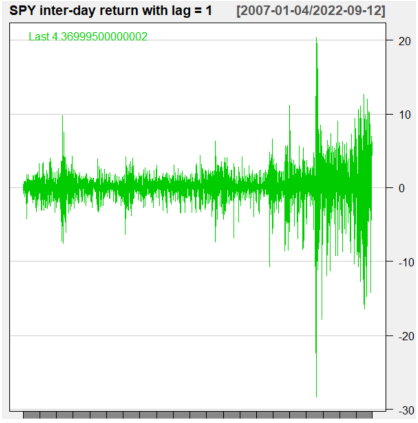

Figure 2: Inter-Day Return of SPY from 2007 to 2022

5.2 Normality Test — Shapiro-Wilk normality test Density Plot and QQ Plot

The Shapiro–Wilk test is a test of normality. It was proposed by Shapiro and Martin Wilk in 1965. Assume a sample x_1, ..., x_n. The hypothesis in which assuing X formed a normally distributed population. The test statistics is calculated by $W_T = \sum_{i=1}^{n}(a_i - x_{(i)})^2 / \sum_{i=1}^{n}(x_i - \bar{x}_i)^2$ where $x_{(i)}$ is the ith-smallest number in the sample X and a_i is calculated by $(m^T V^(- 1)/C$

$$\begin{cases} \text{H}_0\text{: The Distribution follows normally distributed population} \\ \text{H}_A\text{: Otherwise} \end{cases} \tag{1}$$

The result is as the following:

```
        Shapiro-Wilk normality test

data:  percentage
W = 0.86662, p-value < 2.2e-16
```

Figure 3: Shaprio-Wilk Test on the SPY inter-day percentage change

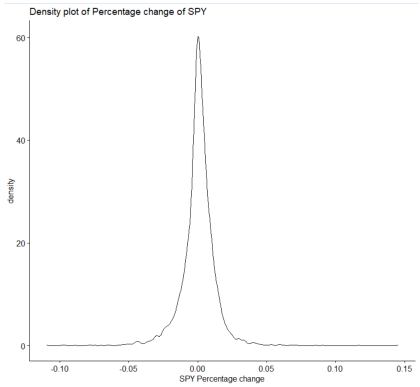

Figure 4: Density Plot of SPY inter-day percentage change

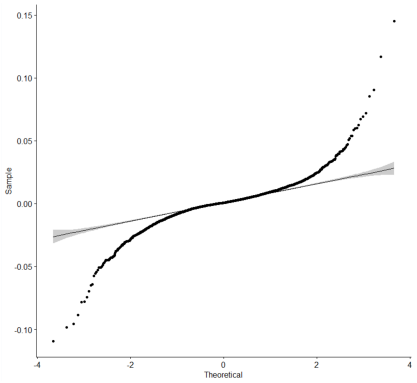

Figure 5: QQ Plot of SPY inter-day percentage change

From the Shaprio-Wilk Test, it is shown that the p-value is very close to 0 and is much smaller than the significance level α. Also, from the QQ Plot, a lot of points, especially those are away from zero, deviate from the theoretical curve. It is suggested that the population does not follow the normal distribution. However, there is a interesting obeservation. From the Density Plot, it is clear that most of the points are concentrated around -1.5% to 1.5%.

5.3 Stationarity of the financial market

To access the auto-correlation, three tests would be tested in the SPY 500, including the ACF Plot, Durbin-Watson Test, and the Breusch-Godfrey Test.

5.3.1 Auto-Correlation Function (ACF) Plot

Autocorrelation occurs when the residuals of a regression model are not independent of each other. In other words, if the price of the SPY 500 is dependent on the price of the previous day(s) and e_i depends on the value of residual e_{i-1}, or sometimes, more than one or more time "lags".

A correlogram, as known as Auto Correlation Function ACF Plot and Auto-correlation plot, could be a visual way to appear a serial relationship in time arrangement information. Serial relationship is where an error at one point in time voyages to a consequent point in time. Correlograms can provide us with a great thought of whether or not sets of data appear auto-correlation.
Due, the correlogram may be used for checking the auto-correlation in a data set. The below result is the ADF Plot for SPY from 2014 to 2022 Sep:

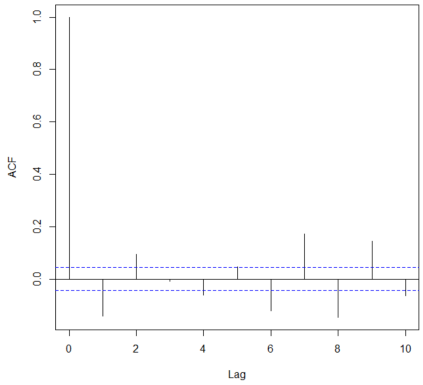

Figure 6: ACF Plot of SPY from 2014 to 2022

5.3.2 Durbin-Watson Test

Another method to measure the autocorrelation of residuals is the Durbin-Watson test. More specifically, it checks the first-order autocorrelation.

Null Hypothesis of Durbin-Watson Test (H_0) : First Order Autocorrelation does not exist.
Alternative Hypothesis of Durbin-Watson Test (H_A) : First Order Autocorrelation exists.

A Autoregressive model: $y_t = B_0 + \beta_0\ y_{t-1} + e_t$ is built if only the first term is focused and it is also expected that the latest would have the strongest effect on the current price. The Test use the following test statistics to determine where the H_0 is rejected or not:

$$DW_T = \sum_{i=2}^{n}(e_i - e_{i-1})^2 / \sum_{i=1}^{n} e_i^2 \ ; 0 \leq \mathrm{DW}_T \leq\ 4$$

If $DW_T < 2$, successive error terms are positively correlated.
If $DW_T > 2$, successive error terms are negatively correlated.

To test for positive autocorrelation at significance α, the test statistic DW_T is compared to lower and upper critical values : [8]

1. If $DW_T <$ lower Critical value, there is significance that the error terms are positively autocorrelated.
2. If $DW_T >$ Upper Critical, there is no significance that the error terms are positively autocorrelated.

3. If lower Critical value $< DW_T <$ Upper Critical, it is inconclusive.

A rule of thumb is that test statistic values in the range of 1.5 to 2.5 are relatively normal. Field(2009) suggests that values under 1 or more than 3 are a definite cause for concern. In other words, if the test statistic value lies between [0,1] U [3,4], it could indicate that very likely there exists auto-correlation. The result is as following:

```
        Durbin-Watson test

data:  model
DW = 1.9783, p-value = 0.6337
alternative hypothesis: true autocorrelation is not 0
```

Figure 7: Durbin-Watson Test

5.3.3 Phillips–Perron (PP) test

Phillips–Perron (PP) test is another test that can be used for evaluating if a time-series data was first-order stationary.[12] it is used in time series analysis to test the null hypothesis that a time series is integrated of order 1. It builds on the Dickey–Fuller test of the null hypothesis p =1, where delta is the first difference operator. Similar to the Durbin-Watson Test, it follows the null hypothesis and the alternative hypothesis as the following:

Null Hypothesis of Phillips-Perron Test (H_0) : Unit root exists and was not stationary.
Alternative Hypothesis of Phillips-Perron Test (H_A) : Unit root does not exist and stationary.

Phillips–Perron test Summary (Jan 2017- Nov 2022) $\alpha = 0.01$		
Tickers	Test-Statistics	P-value
SPY	-2.0588	0.5690
HSI	-1.52893	0.8190
FTSE	-2.5836	0.2874
000001.SS	-2.233	0.4710
GDAXI	-2.54336	0.30667
N225	-2.715	0.2295
BTC-USD	-1.614	0.78545
DX-Y.NYB	-0.110	0.99
CL=F	-2.11	0.5387
NG=F	-2.39	0.382
GC=F	-1.06	0.935

From the above, it is clear that the p-value of the above major indexes are much higher than the significance level(0.01), due the null hypothesis is not rejected, indicating that unit root may exist and the price movement is not stationary.

5.3.4 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test

From the above, it is shown that the DW Test Statistics value is very close to 2 and the p-value is much bigger than significance level $\alpha = 0.05$. Therefore, accept H_0, and conclude that there is very likely that correlation does not exists for time-lag = 1.

Kwiatkowski–Phillips–Schmidt–Shin Summary (Jan 2017- Nov 2022) $\alpha = 0.01$	
Tickers	Test-Statistics
SPY	0.506
HSI	0.453
FTSE	0.7502
000001.SS	0.454
GDAXI	0.435
N225	0.521
BTC-USD	0.5321
CL=F	1.123
NG=F	1.125
GC=F	0.793

From the above, it is clear that the p-value of the above major indexes are much higher than the significance level(0.01), due the null hypothesis is not rejected, indicating that unit root may exist and the price movement is not stationary.

5.3.5 Augmented Dickey-Fuller (ADF) Test

Augmented Dickey-Fuller (ADF) Test is a test that test the null hypothesis that a unit root is display in a time series. The alternative theory is diverse depending on which form of the test is utilized, but is as a rule stationary or trend-stationary. It is an increased adaptation of the Dickey–Fuller test for a bigger and more complicated set of time arrangement models. The increased Dickey–Fuller (ADF) measurement, utilized within the test, may be a negative number. The more negative it is, the more grounded the dismissal of the theory that there's a unit root at a few level of confidence.

Augmented Dickey-Fuller test Summary (Jan 2017- Nov 2022) $\alpha = 0.01$		
Tickers	Test-Statistics	P-value
SPY	-2.200	0.48
HSI	-1.679	0.759
FTSE	-2.6387	0.262
000001.SS	-2.3295	0.417
GDAXI	-2.687	0.241
N225	-2.953	0.145
DX-Y.NYB	-0.0274	0.993
CL=F	-1.831	0.689
NG=F	-2.283	0.443
GC=F	-0.736	0.970

From the above, it is clear that the p-value of the above major indexes are much higher than the significance level(0.01), due the null hypothesis is not rejected, indicating that unit root may exist and the price movement is not stationary.

5.4 Volatility Estimation Using GARCH Models

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) Models is used to estimate the volatility. Institutional Traders price assets and detect which one potentially provide the best return with the adjustment of portfolio allocation and risk management.
There are a variety of forms of ARCH models, but in the following, the test would be tested on the "Garch (1,1)" Model.
After the modelling, it is first tested against the VIX (SP500 short futures) by the modelled SPX (SP500 ETF) price ignore the ETF loss, and assuming the SPX has the same movement as SP500. The result is as the following:

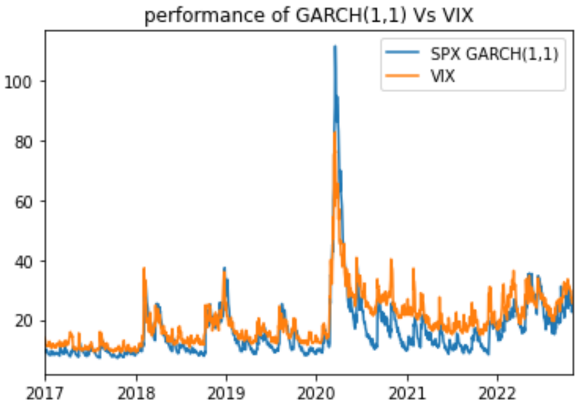

Figure 8: GARCH Performance

From the above, it is shown that the GARCH(1,1) has a roughly the same movement as the VIX price, showing a good sign of the volatility modelling.
Then, the GARCH is used to test for the market risk of different regions. US market, Japanese market (Nikkei 225 Index), Hong Kong Market (Heng Seng Index) and French Market (Cotation Assistée en Continu 40 Index) would be compared to see the overall market risk from 2006 to 2022.

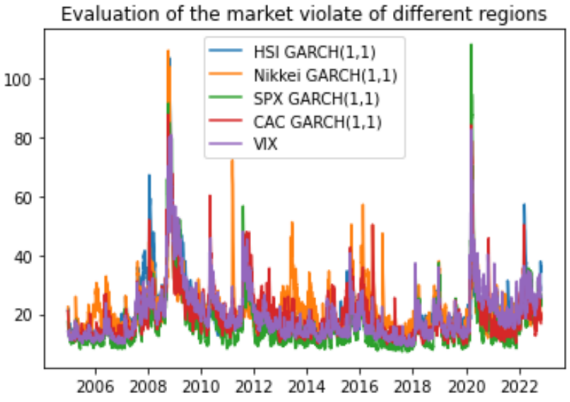

Figure 9: Market Volatility of different regions modeled by GARCH

From the above, it is shown that from the beginning of 2000, 2011, 2014, Japanese market, in general, has the highest market risk compared to the other markets. It be attributed to the fact of the "Japanese asset price bubble". In 2008-2009 and 2020, all market has a extremely volatility. This may be attributed to the financial crisis and the covid-19 pandemic. These are in line with the assumptions and proves the accuracy of the GARCH model.

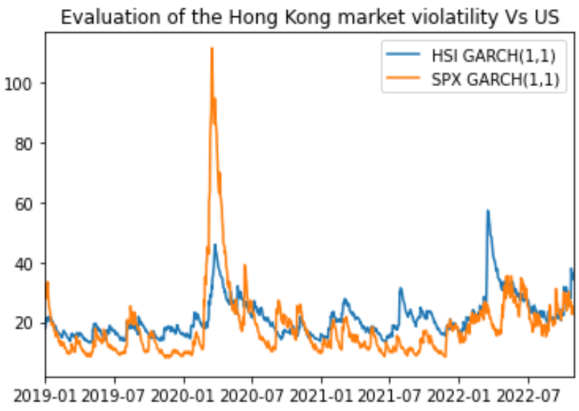

Figure 10: Market Volatility of Hong Kong Stock Market modeled by GARCH

Finally, I want to evaluate the volatility of the HK market compared with the US market in the recent years. From 2020-2022, the volatility of the HK market is generally higher than the US. However, the spread in the value is not as high as what I expected. Only the latest 2 months has a major different in the value, possibly because of the sell-offs of the Chinese stock in late October 2022. It may also be concluded that the global economic recession is the main drive of the crash of the Hong Kong market in the recent three years. Further research is needed to prove the above statement.

5.5 Option Pricing and open premium — Black Scholes Model

Derivatives market consist of different instruments and players. Stock option is one of the most popular ones. There are mostly two major types of options. The first one is "call options" which bets a particular stock will increase in value before the the expiration date, or puts, betting a stock will decrease in value before the expiration date. But of course, there are different strategies that options can take play such as vertical call/put, spread call/put, butterfly, straddle, etc. Our research will solely take place in either call of put. The Black Scholes model, also known as the Black-Scholes-Merton (BSM), is considered to be one of the best models to determine the fair prices of stock options. The model is followed by the formula: $C{=}SN(d_1)\text{-}Ke^{-rt}N(d_2)$ where C= Fair Price of the call option, S= Current market stock price, K=strike price, r=interest-free rate, t= time to expire, N= normal distribution. In this section, we would test how fit the Black Scholes Model to the real-time trading scenario, by comparing the fair price and the bid-ask price. The result is as the following:

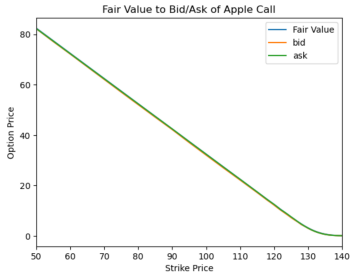

Figure 11: Fair Value of Apple Inc Call

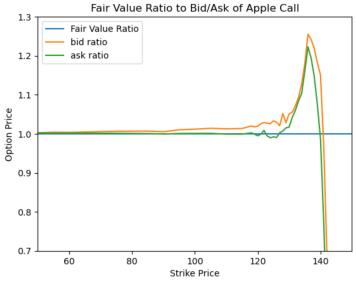

Figure 12: Fair Value of Apple Inc Call

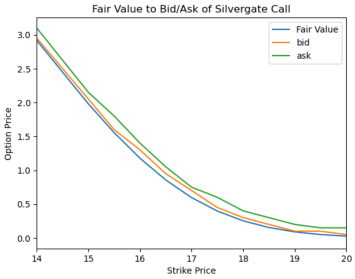

Figure 13: Example of a Risky Stock (silvergate Capital)

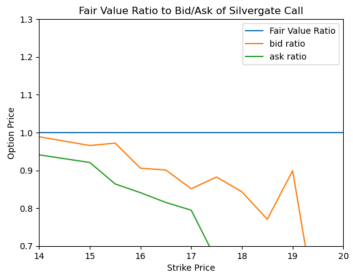

Figure 14: Example of a Risky Stock (silvergate Capital)

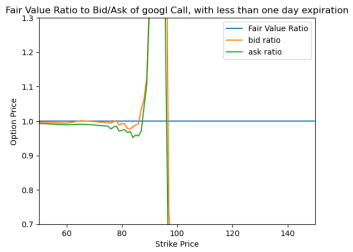

Figure 15: Example of almost expired calls

From the above, it is clear that for the stocks with stable standard deviation and high liquidity, such as apple, Costco, Microsoft, and Google, the bid and ask price of the options roughly follows the Black Scholes Model, omitting the insufficient calculation error in interest free rate and time to expire. The fair price ratio approximates to 1 for most of the in-the-money call and slightly deviate when the expiration time is short and for out-of-money call. However, for some small cap stocks or stocks with high standard deviation, the bid and the ask price can deviate a lot from each other and that of the fair value. Also, some calls that near to expiration can have significant different bid ask price than the fair value. The difference in the fair price and bid/ask price may be attributed to the preference of the investors where some non-factors are not considered in the model such as the "earning report release" or potential movement driven by possible company action.

5.6 Monte Carlos Simulation — GBM

Monte Carlo simulation, also known as Monte Carlo experiments, is a mathematical model which is used to predict the probability of a variety of outcomes by computational algorithms that utilizes random sampling to obtain numerical results for different possible outcomes. Monte Carlo methods varies from different fields of study, like physics, mathematics, and computer science. But it usually requires: definition of a domain of possible inputs, randomly generated inputs from distribution over the domain, and deterministic computation on the inputs and aggregation of the results. [6] In the financial market study, by assuming the market is perfectly efficient and history repeat itself, Monte Carlo simulations is commonly used to explain the risk and uncertainty in prediction for forecast testing, by assigning multiple values to an uncertain variable to achieve multiple results and then averaging the results to obtain an estimate. In this section, we would test the geometric Brownian motion(GBE) which follows the Markov process which we assumes the overall trend follows a random walk and follows the weak form of the efficient market hypothesis (EMH)[3]. The GBE model follows :$\Delta S = S(\mu\Delta t + \sigma\epsilon\sqrt{\Delta t})$ The result is as the following: For the long time frame, the Monte Carlos Simulation fails due to the

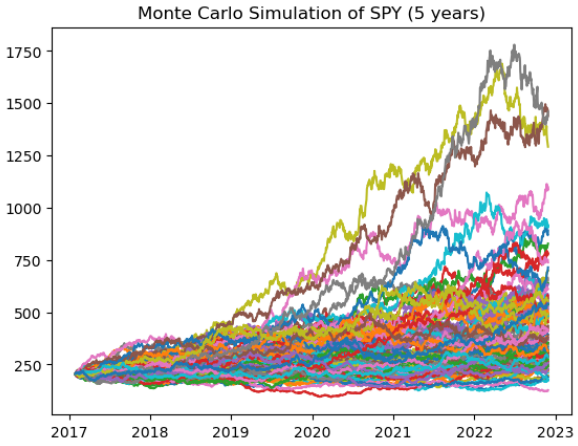

Figure 16: Monte Carlos of SPY in 5years)

more uncertainly in the long run and compound effect of the error which makes the random simulation varies a lot from the actual price curve. In this regard, a rolling window training is applied to the Monte Carlo Algorithm to improve the performance.

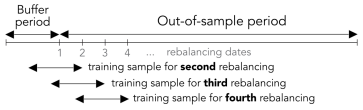

Figure 17: Demonstration of the rolling Windows)

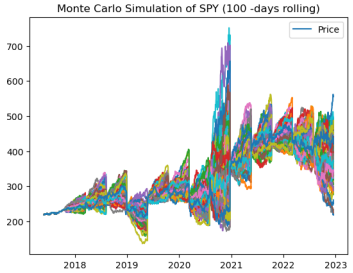

Figure 18: Monte Carlos in rolling windows of 100 days)

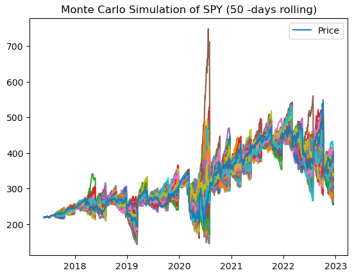

Figure 19: Monte Carlos in rolling windows of 50 days)

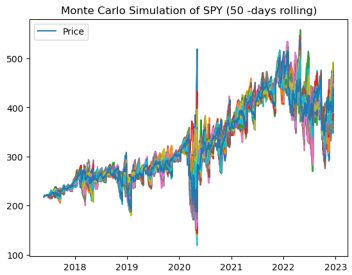

Figure 20: Monte Carlos in rolling windows of 20 days

From the above, it is clear that the short the rolling time frame is, the better fitness of the Monte Carlos simulation is with the actual price curve. The Monte Carlos Simulation could be a feasible way to forecast the price movement a particular stock.

6 Back-Testing Results

6.1 Random Flip-coin Strategy

There is a belief that most of the retail investors suffer loss in the stock market because of the unsystematic system. Especially to some intra-day traders, the retail investors usually have the illusion that the stock rises today will more likely to falls tomorrow due to the lose of momentum. However, suggested by many research, it is more likely a psychological belief or a subjectivity. Due to the randomness walk nature of the stock market and the auto-correlation, the subjectivity and trading randomly is regarded as a bad strategy in the long run.
Therefore, in the following, Random Flip-coin strategy is created to modify the trading behavior of the retail investors: Denote the long trade and short trade as a random discrete binary variable x where:

$$x = \begin{cases} 1 & ; \, if \; it \; is \; a \; LONG \; Trade \\ 0 & ; \, if \; it \; is \; a \; SHORT \; Trade \end{cases}$$

where $P\{\ \mathrm{X} = 0\ \} = 0.5$ and $P\{\ \mathrm{X} = 1\ \} = 0.5$
For simplicity, the strategy is tested on the daily SPY candle chart data from 1 Jan 2017 to 31 Aug 2022 and the return is calculated by $Return = Close - Open$. Therefore, it is expected return of the strategy is Calculated recursively by:

$$E(R_t) = \begin{cases} 1 & ; \, if \; t \; = \; 0 \\ E(R_{t-1}) * (Close_t / Open_t) & ; \, if \; x_t \; = \; 1 \\ E(R_{t-1}) * (Open_t / Close_t) & ; \, if \; x_t \; = \; 0 \end{cases}$$

The result is shown as the following:

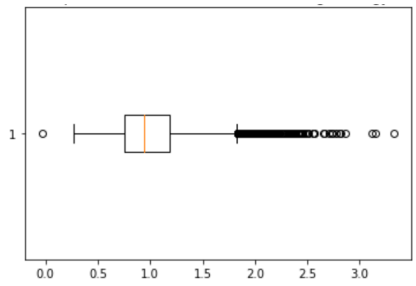

Figure 21: Expected Return of Random Flip-coin over 5 years on SPY

The expected return is around 1 which is the same as the theoretical mean of the binary variable and the random movement nature of the stock market. However, the above simulation omitted the effect of the price spread and the commission fee. Repeat the above algorithm and back-test with different degree of price adjustment:

$$E(R_t) = \begin{cases} 1 & ; \, if \; t \; = \; 0 \\ E(R_{t-1}) * (Close_t / Open_t) * priceAdjustment & ; \, if \; x_t \; = \; 1 \\ E(R_{t-1}) * (Open_t / Close_t) * priceAdjustment & ; \, if \; x_t \; = \; 0 \end{cases}$$

The result is as following:

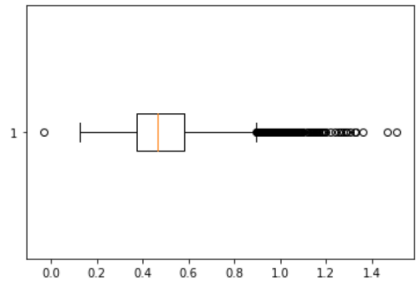

Figure 22: Expected Return of Random Flip-coin with price adjustment =0.0005

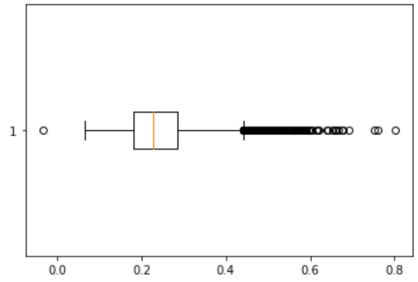

Figure 23: Expected Return of Random Flip-coin with price adjustment =0.001

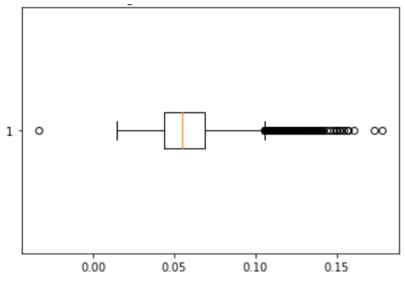

Figure 24: Expected Return of Random Flip-coin with price adjustment =0.002

With the Price Adjustment added. The performance of the strategy greatly decreases with time due to the compounding effect of the price adjustment, which inline with our expected results, even the expected value of the distribution should be one. Theoretically, it is also suggested that more frequent the trade happens. Our strategy is tested on the basis of trading daily and using ordinary stocks only. However, it expected that higher price adjustment and more loss in the commission fee and spread for the intra-day traders or using different derivative products such as options and futures. Implication: Random Flip-coin Strategy fails. It is a must to find a profitable strategy with expected value of each trade greater than 1. Therefore, a more advanced trading strategy is required in order to generate a huge profit in long run.

6.2 Deep Learning — Multi-layer perceptron

From the previous Random Flip-Coin Strategy, we have proved that without clear decision rule, investor will very likely suffer a loss due to the compound effect of commission fee and short interest. However, what else can we do with the randomness? The first trail is the use of Multilayer Perceptron (MLP), primarily based on the close price, relative strength index, moving convergence and divergence indicator and the boiling band of the stock. The price is this model is calculate by the percentage change of the previous N days in different time interval (15mins, 60 mins and 1 day).The model is evaluate and retrained. The MLP network is as the following:

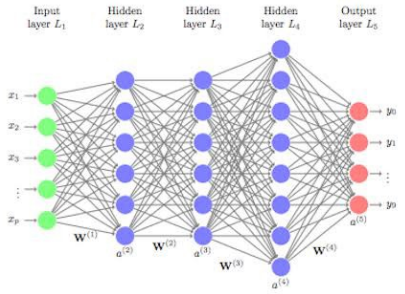

Figure 25: Example of the MLP architecture

MLP Summary (Test:SEP 2022- Nov 2022)			
Model	Indicators	Accuracy	Loss
Linear(40,25,13)	Price	0.45	0.736
Relu(20,10,5)	Price	0.455	0.75
Relu(40,25,13)	Price	0.492	0.759
Relu+RSI(41,25,5)	RSI+Price	0.515	0.680
Relu+RSI(20,13,5)	RSI+Price	0.51	0.6829
Relu+RSI+MACD(42,15,7)	RSI+Price+MACD	0.54	0.657
Relu+RSI+MACD(20,13,5)	RSI+Price+MACD	0.535	0.689
Relu+RSI+MACD+BB(43,13,5)	RSI+Price+MACD+BB	0.56	0.645
Relu+RSI+MACD+BB(43,13,5)	RSI+Price+MACD+BB(60mins)	0.585	0.657
Relu+RSI+MACD+BB(43,13,5)	RSI+Price+MACD+BB (15mins)	0.576	0.665

From the above, we can see that the model generally perform differently ranging from the time interval used, indicators, asset price. More the information the model gets, slightly better the performance it has. Shorter Time interval, the better the performance it has under our strategy, possible attributable to the sentiment and momentum driven without the effect of the market news. However, the validation accuracy of the testing data-set is still below 0.6 which is not significantly higher than the threshold value (0.65) for the binary classifier MLP. The reason may be solely depends on the price movement does not give it sufficient information and very sensitive to the random noise of the market movement.

6.3 Asset Allocation Strategy — Markowitz Portfolio Optimization

The idea of Markowitz's mean-variance portfolio (MVP), which is also known as Modern Portfolio Theory (MPT), is based on the compromise of the expected return and the risk of a portfolio. Despite the assumptions, such as investors are risk averse and increase consumption, investors and rational and the market contains of perfect information, which is always not true in reality, Markowitz Portfolio optimization is one of the most popular portfolio optimization method due to its simplicity, but powerful performance. The model relies on two important psychological behavior. Under the rational assumption: 1. For the same risk level, investors would prefer the stock with higher return. 2. For the same return, investors would prefer stock with lower risk. The risk and return can be qualified as standard deviation and expected return ($w^T\mu$ and $w^T\sum w$), where w is denoted as weights. The best portfolio with the highest return on the same risk level is called "Efficient Frontier". There different forms of portfolio optimization model that deviate from the MVP such as, Maximum Sharpe ratio portfolio (MSRP) and Maximum Sharpe ratio portfolio (MSRP). General Mathematically speaking, the goal is to:

1. Maximize $w^T\mu \leq \lambda w^T\sum w$ (MVP)
   OR minimize $w^T\sum w$ (GMVP)
   OR maximize $(w^T\mu - \tau_f)/\sqrt{w^T\sum w}$ (MSRP)
2. budget constraints $w^T = 1$ (i.e. No Borrowing)
3. $w >= 0$ (i.e. Long Only Strategy)
4. $||w|| \leq$ u (i.e. max weight of a position)
5. $||w||_1 \leq \tau$ (i.e. leverage)
6. $||w||_0 \leq K$ (i.e. sparsity)
7. $\sigma_{portfolio} = \sqrt{\sum w_i\sigma_i}$

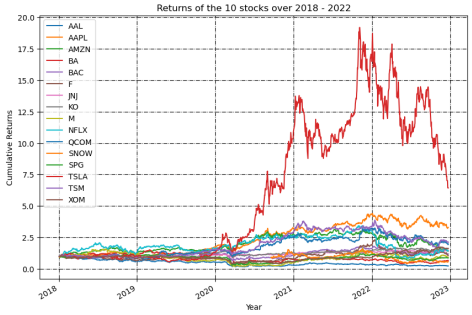

Figure 26: Example of Cumulative returns of Stocks

Markowitz Optimization Summary (Jan 2018- Dec 2022) (10 trails of 15 random stocks from SPY)			
Model	Combination of tickers and weights (nearest integer)	Annualized Returns	Annualized risk
Sharpe ratio	[AAPL,AMZN,JNJ,NFLX,TSLA,TSM]=[14,17,14,17,14]	0.27	0.253
Min risk	[AMZN,JNJ,KO,NLFX,TSM,XOM]=[5,42,38,1,7,5]	0.12	0.15
Sharpe ratio	[JNJ,KO,TSLA,TSM,XOM,ZM]=[6,31,41,11,8,2]	0.32	0.32
Min risk	[JNJ, KO, TSM, XOM, ZM]=[43,35,5,7,7]	0.12	0.17
Sharpe ratio	[A,TSM,XOM,ZM]=[49,22,16,10]	0.19	0.25
Min risk	[A,BCA,F,GOOGL,SPG,TSM,XOM,Z,ZM]=[31,3,2,15,2,8,30,8]	0.34	0.40
Sharpe ratio	[A,NET,NVDA,OXY]=[43,32,16,8]	0.19	0.25
Min risk	[A,BCA,F,GOOGL,SPG,TCOM]=[43,10,6,28,5,6]	0.15	0.25
Sharpe ratio	[NET,NVDA,ORCL,OXY,TGT]=[26,13,14,7,39]	0.32	0.36
Min risk	[F,ICE,ORCL,SPG,TCOM,TGT]=[2,50,22,3,8,16]	0.14	0.22

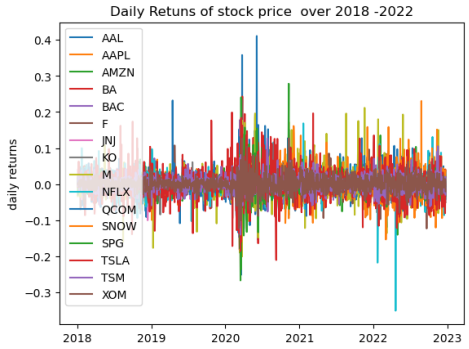

Figure 27: Example of Return of assets over the period

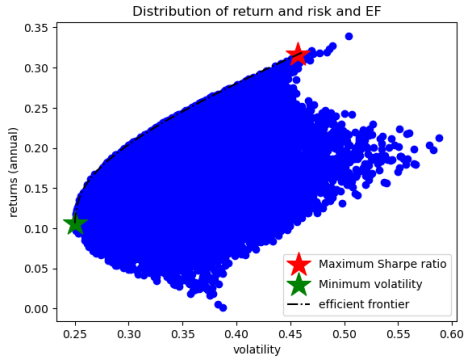

Figure 28: Example of Efficient frontier

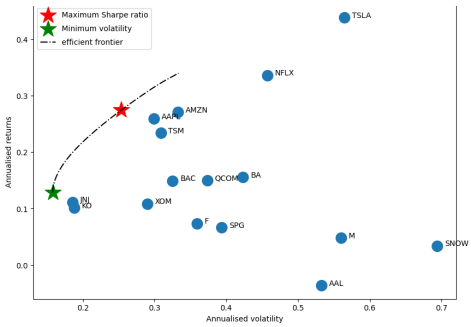

Figure 29: Example of Selected Combination and efficient frontier

6.4 Long Short-Term Memory (LSTM) — Price Prediction

LSTM is a type of neural network that similar to the MLP. But compared to MLP, LSTM, which is also a type of RNN model, has higher compatibility to deal with the sequence data. In the previous MLP model, it is clear that the MLP fails to predict the stock return direction due to the time-series nature. LSTM model include a 'memory cell' that can maintain information in memory which lets them learn longer-term dependencies [10]. In general, When data is processed 1 step through LSTM, it behaves roughly the same as MLP given both have same network structure. However, when data is processed with more than 1 time unit, it is expected to perform much better than the MLP network. In this section, LSTM prediction capacity is accessed against SPY price.

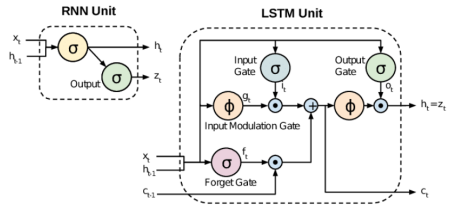

Figure 30: Example of a LSTM network

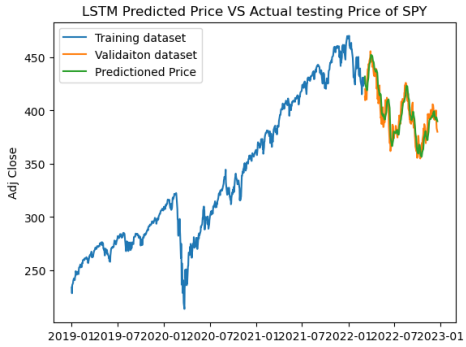

Figure 31: Example of a one of the result LSTM

LSTM Summary (Jan 2019- Dec 2022), Epoch = 3	
Network Structure	MSE(Validation)
LSTM 100 (return), LSTM 50 (non-return),1(linear)	6.12
LSTM 100 (return), LSTM 50 (non-return), 25(linear),13(linear),1(linear)	2.95
LSTM 100 (return), LSTM 50 (non-return), Dropout(0.25)25(linear),13(linear),1(linear)	0.98
LSTM 100 (return), LSTM 50 (non-return), 25(relu),13(relu),1(linear)	3.9
LSTM 200 (return), LSTM 100 (non-return), 50(relu),25(relu),1(linear)	4.15

From the trial and error result, it is shown that the LSTM network has a high capability to learn and predict the stock price accurately. The LSTM outperforms MLP network in a significant way. the lowest MSE is 0.98 which is back-tested with the validation data from Jul 2022 to Dec 2022. It is also found that changing the time stamp will have significant impact on the performance .

6.5 Pair Trading — K-means

6.5.1 Brief Summary

Algorithm of the K-means pairs trading:

1. Substract the adjusted close price of the stocks.
2. Data Prepossessing
3. Carry out the K-means clustering algorithm and determine the best "K" value using elbow method and silhouette method
4. Sort the Cointegrated pairs by either augmented Dicky-Fuller (ADF) test and Johansen test to check if the spread of pairs are stationary by the p-value.
5. Sort the Correlated pairs by Pearson test to check if they co-move or move independently
6. Calculate the Hurst Exponent. Sort the pairs with Hurst Exponent < 0.4, which is calculated by $E[(R_n)/S(n)] = Cn^H$, which a measure of long-term memory of time series, indicating the mean-reverting behavior.
7. Calculate the Half-Life, which is calculated by and indicates how long the spread typically takes to revert back to the mean. Sort the pairs with $10 <$ "half-life" < 50,
8. Sort the pairs with mean-cross > 12.

Inputs tickers: stocks under SP 500
Time period: Jan 2020- Nov 2022
Models: K-means with absolute decision rules

6.5.2 K-means Clustering

After the K-means clustering, elbow method and silhouette method is taken with the respect to K=5 and K=4. In this case we would use K=4 and result is as the following:

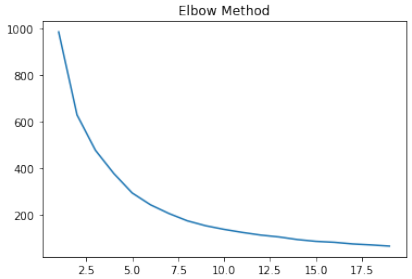

Figure 32: Elbow Method

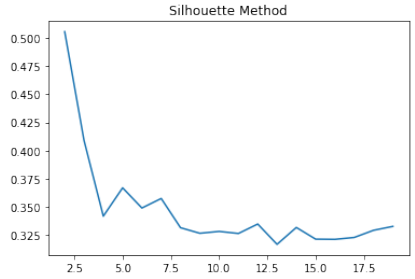

Figure 33: Silhouette Method

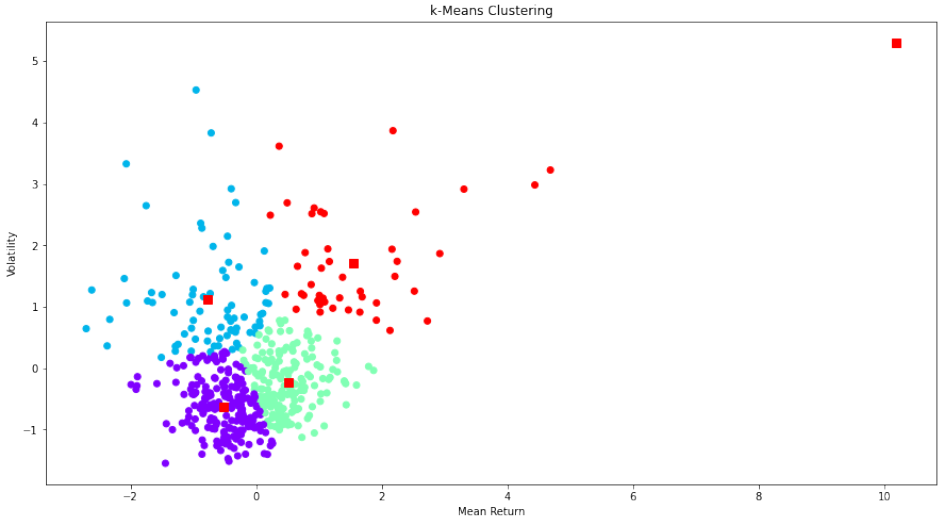

Figure 34: K-means clustering results

6.5.3 Additional Decision Rule

Pairs Selection Summary	
State	No. Paris
Filter	38343
Cointegration	1110
Correlation	471
Hurst Exponent	190
Half Life	120

Compared to the traditional Way of find trading pairs inside the same sector or by calculating the absolute statistical value, the use of K-means Clustering can be a innovative to find pairs. The number of pairs found in the first few filters is far more than the traditional research study. Also, by applying these decision rule, the paiss should be of high liquidity, cointegrated, correlated, mean-reversion behavior, which makes them suitable for the pairs trading. In the next subsection, the trading performance of the trading pairs is evaluated.

6.5.4 Trading Rule

After getting the pairs, different pairs trading rule can be applied. This section, the mean-reversion strategy is used which takes advantage of the pairs when converge and divergence. The assumption behind the strategy is that the pairs should have similar behavior after the pairs selection and will converge to the normal level after sufficient time. The normal level can be calculated using either OLS regression or Z-score.
From the above chart, it is clear that the investor can take profit when the price deviate from the normal price by shorting/longing (for example red line = $\mu + \pm\ 2\ \sigma$). Then by closing the position at

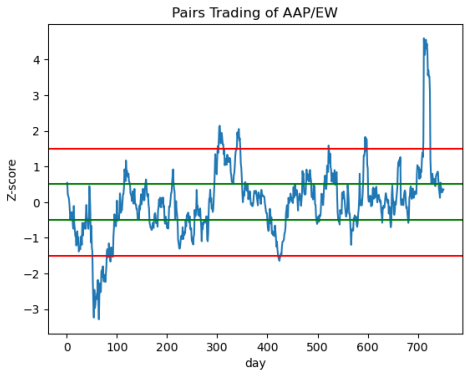

Figure 35: Example of AAP/EW

the normal range (for example the green line = $\mu + \pm\, 0.5\ \sigma$. This strategies can further optimized by constructing a portfolio of all pairs got from the K-means results by Kelly betting [16] or Minimum Variance, our statistical arbitrage opportunity.

7 Limitation and Potential Improvements

7.1 Selections of Financial instruments

There are different instruments in the financial market other than stock, such as bond, stock options, Cryptocurrency, futures, warrant, cfd, and swap. This research primarily research on pure ordinary stock. Different combination of securities may perform better due to the less efficient market.

7.2 Option Pricing

This research primarily focus on the day time interval. However, the underlining stock price, risk-free rate and the time to expire can change from minute to minute. The Black Scholes Model use omit the effect of dividend. Since the dividend indicates cash outflow from company to investors, while the option holder will not receive dividend, the dividend payout should decrease the company book value and due decrease the fair value of the option. [11]Also, the only the single call option is accessed. However, in reality, the actual option strategies usually consists of put options together to form different option strategies such as straddles, spread and vertical.

7.3 Markowitz Portfolio Optimization

Similar to the Option Pricing, this research omit the effect of dividend yield rate. Stocks with high dividend yield, such as XOM, OXY, T, VZ should have higher return, but is not included in the calculation of the annualized return. Also, this research primarily focus on the biggest 100 stock in SP500, however, there are more research indicate that there could be much more promising results if we include the T-bond, FX, etc.

7.4 LSTM prediction

In this section, price is predicted using LSTM model. Nevertheless, the LSTM model is commonly also combined with the use of portfolio optimization. [1]

7.5 K-means Paris Trading

In this section, only pairs with two stocks are accessed. However, in the actual trading world, it is common to use more than 10 securities in one single pair. Also, with the pairs trading, it is also feasible to combine the pairs results with the portfolio optimization technique by randomly assigning weights to the pairs. Also, Some research shown that pairs trading can be more profitable by statistical arbitrage or constructing a mean reverting portfolio.

References

[1] Mike Bernico. "Deep Learning Quick Reference: Useful hacks for training and optimizing deep neural networks with TensorFlow and Keras". In: (2018).

[2] Imad Dabbura. "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks". In: *Towards Data Science* (2018). URL: https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a.

[3] KATRINA MUNICHIELLO DAVID R. HARPER GORDON SCOTT. *How to Use Monte Carlo Simulation With GBM*. Investopia, 2019.

[4] Sebastian Harder. *The Efficient Market Hypothesis and its Application to Stock Markets*. GRIN Verlag, 2010. ISBN: 9783640743766.

[5] Forest Reinhardt José B. Alvarez and Natalie Kindred. *CME Group in 2019*. Harvard Business School, 2019. URL: https://www.hbs.edu/faculty/Pages/item.aspx?num=57440.

[6] Rohan Joseph. *The house always wins : Monte Carlo Simulation*. IBM Cloud Education, 2017.

[7] Robert Kissell. *Algorithmic Trading Methods Applications Using Advanced Statistics, Optimization, and Machine Learning Techniques*. Academic Press, 2020. ISBN: 9780128156315. URL: https://www.hbs.edu/faculty/Pages/item.aspx?num=57440.

[8] David C. Harris Maxwell L. King. *The Application of the Durbin-Watson Test to the Dynamic Regression Model*. Department of Econometrics, Monash University, 1995.

[9] Pradeep Pujari Mohit Sewak Md. Rezaul Karim. *Practical Convolutional Neural Networks: Implement advanced deep learning models using Python*. Packt Publishing, 2018. ISBN: 9781788392303.

[10] Ashutosh Tripathi NLP. *WHAT IS THE MAIN DIFFERENCE BETWEEN RNN AND LSTM — NLP — RNN VS LSTM*. Data Science Duniya, 2021.

[11] *Option Pricing Theory and Models*. New York University, 2007.

[12] Peter C. B. Phillips and Pierre Perron. *Testing for a Unit Root in Time Series Regression*. Oxford University Press, 1988.

[13] Federal Reserves. "The Fed Explained What the Central Bank Does". In: *Federal reserve system publications* (2021). DOI: https://doi.org/10.17016/0199-9729.11.

[14] I.H. Sarker. "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions". In: *SN Computer Science* (2021). DOI: https://doi.org/10.1007/s42979-021-00815-1.

[15] Nunzio Tartaglia. *THE COMPREHENSIVE INTRODUCTION TO PAIRS TRADING*. Hudson Thames, 2020.

[16] CFI Team. *Kelly Criterion*. Corporate Finance Institute, 2021.

List of Figures