## The University of Hong Kong Department of Statistics and Actuarial Science STAT3609 The Statistics of Investment Risk

## 2020-21 Semester 1

## **Group Project**

# Group 2 < Candlestick Analysis: Prediction of Hong Kong and Korean Stock Market Trend >



#### Members:

| <u>Name</u> | <u>UID</u> |
|-------------|------------|
| LEE ***     | ****       |
| CHOI ***    | ****       |
| YEUNG ***   | ****       |
| WONG ***    | ****       |

Declaration: The work has not been used for any other courses by any of the members.

# I. Background

#### I.1. Introduction

Candlestick pattern, one of the technical analysis tools, has been taking reference by investors and traders to determine the price directions. In this project, candlestick patterns are explored whether they are applicable to the real financial market, and hence discover any difference in the patterns' significance between Hong Kong and Korean financial markets.

#### I.2. Candlestick

A candlestick pattern is one of the technical indicators in technical analysis. First of all, technical analysis is a trading discipline adopted to explore or evaluate investments and identify trading opportunities by analyzing historical trading data of any securities using technical indicators (HAYES, 2020).

There are lots of technical indicators such as moving average, support and resistance level. Candlestick patterns are also commonly used because of its simplicity in concept and application. In this section, symbols, meanings and usages of candlesticks will be introduced, followed by specific candlestick patterns.

#### I.2.1. Basics

A candlestick is a kind of price chart which displays the high, low, opening and closing prices of a stock or security for a specific period (HAYES, 2020). Figure I.1.1.1 below takes Period=1 as an example to illustrate its composition.

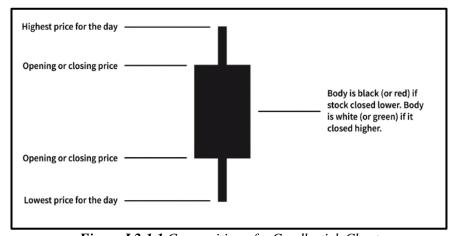
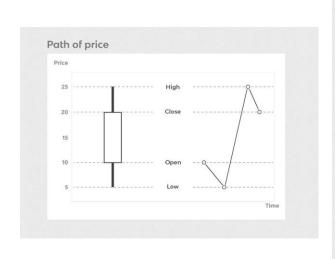


Figure 1.2.1.1 Composition of a Candlestick Chart

The head of a candlestick is the highest price for the day, while the bottom is the lowest price. The rectangular box is called a body. The head and the bottom of a body are the opening or closing price. Their actual positions are determined by whether the day (Period = 1) is bullish or bearish. If it is bullish, the upper one will be closing price and the lower one will be opening price, vice versa. The condition of the day is also shown in a candlestick. The body is empty when the market is bullish, while it is filled when the market is bearish. Therefore, many summary statistics can be obtained by simply looking at this little candlestick.

To further elaborate, Figure I.2.1.2 (Rolf, 2019) illustrates a simplified version of the path of the security on the corresponding day. Nevertheless, when combining a series of candlesticks, a path with a longer period of the stock can be explained, like in Figure I.2.1.3 (Rolf, 2019).



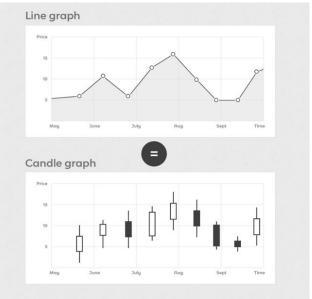


Figure I.2.1.2 Path of stock price

Figure I.2.1.3 Connected candlestick ling graph

#### I.2.2. Candlestick Patterns

A combination of numbers of candlestick charts is called a candlestick pattern. Candlestick patterns make candlestick itself more useful than only the open-high, low-close bars or simple path of the price (Farley, 2020). By observing candlesticks build patterns, investors can predict the future market trend or momentum. For example, the price of the stock is likely to rise for a period of time if a candlestick pattern Inverted Hammer is found.

There are a lot of candlestick patterns that can be used to make profitable financial decisions, including long the securities, short or even short sell the securities, buy some derivatives, etc.

Some investors believe candlestick patterns are a reliable indicator. However, are they, actually? From the nature of future prediction using historical data, there will not be any indicators with 100% certainty, otherwise, the market would be weak form efficient. Therefore, this project is an attempt to test the accuracy and significance of the prediction, and applicability of different patterns in different financial markets.

#### I.2. Indices

## I.2.1.Hang Seng Index

Hang Seng Index is a market-value-weighted index record and monitors daily changes of the largest companies of the Hong Kong stock market. It is the main indicator of the overall market performance in Hong Kong. Its constituent stocks include Tencent, Swire and MTR.

#### I.2.2. KOSPI 50 Index

Korea Composite Stock Price Index 50 is a market-value-weighted index and representative stock market index of South Korea. KOSPI has over 880 components. Top 3 stocks are Samsung Electronics, SK Hynix and LG Chem. KOSPI 50 is chosen over KOSPI 200 to match with the number of stocks with HSI. The advantage is its price change has little discrepancy with KOSPI200, a more commonly used index and the weight of top 50 stocks among 200 stocks is high.

# II. Methodology

Before explaining each step, basic terms in Python would be explained in advance. The application was done by Python coding on Jupyter Notebook. Jupyter Notebook is a web application often used in data science since it allows live code running (refer to the html code file attached). *Pandas* is a Python library for data analysis, and it uses *dataframe* as the primary 2-dimensional data structure with labelled rows and columns.

#### II.1. Data Collection

Daily open, high, low, and close (ohlc) price and the corresponding date was required for the project. Any open-source files in any format, such as csv and json could be processed in Python. Major financial websites such as Yahoo finance, investing.com, Wall Street Journal, etc. would be the primary target for data searching. The preferable time range was around one decade.

## II.2. Data Preparation

From the raw data collected from the external sources, data needed to be cleansed and modified for further process. First, two basic price dataframes of indices were constructed from the data files. Then, unnecessary columns should be dropped and the data frames should be unified to descending order by the date column.

#### II.2.1. Candlestick Chart

Before further modification, candlestick charts would be drawn with *mpl\_finance* library to show the general price change.

## II.2.2. Candlestick Pattern Recognition

The library *TA-Lib* supports diverse technical analysing functions of financial market data, such as volume indicators, momentum indicators, etc (mrjbq7, n.d.). Among those, pattern recognition functions would be used in this project. There are 61 defined pattern functions in TA-Lib which take ohcl prices as input and return positive or negative integers. A positive integer indicates bullish pattern, a negative integer indicates bearish pattern, and null value indicates no such pattern is found. The results of one pattern would be recorded as one column of a new dataframe, so that finally there are 61 columns and every row indicates one day.

Furthermore, different t-test would be done for bullish and bearish patterns. Therefore, by the sign of values in the dataframe, the dataframe would be separated into two datasets, one for bullish patterns and another for the bearish patterns. Then, to analyse only the meaningful patterns, the ones with occurrences of less than 25 were dropped out from each dataset.

Yet, the integers from TA-Lib are rather meaningful in this project. Simple return defined as the equation below would be used in the t-test in Section II.3.

$$R = \frac{P_1 - P_0}{P_0} = \frac{P_1}{P_0} - 1$$

Therefore, simple return values should replace the integers or null values from the previous dataframe.

## II.3. T-test

One sample t-test (one-tailed) was used to determine whether the average simple return of a specific candlestick pattern is statistically different from a hypothesized population mean.

## II.3.1 Assumption

There are mainly four assumptions for the one sample t-test. First, the dependent variables (in this project, simple returns) are continuous. Second, the data are independent of each other. Also, the dependent variables should be approximately normally distributed. We talk about the one-sample t-test only requiring approximately normal data because it is quite robust to violations of normality, meaning that the assumption can be a little violated and still provide valid results. Since the sample size is limited to be 25 or above, this requirement is considered to be satisfied. Lastly, there should be no significant outliers, which can reduce the accuracy of the test.

To ensure the validity of normality assumption, a quantile-quantile (Q-Q) plot which shows the distribution of the data against the expected normal distribution was used. For normally distributed data, observations should lie approximately on a straight line. If the data is not normal, the points form a curve that deviates markedly from a straight line. Figure II.3.1.1 and Figure II.3.1.2 show the Q-Q plots of 5-day simple returns of HSI and KOSPI50 respectively and the plots are generated using the python library. All the patterns approximately lie on a straight line which indicates the normality assumption is valid.

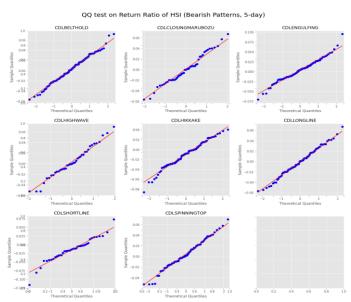


Figure II.3.1.1 Q-Q plot of returns for HSI bearish patterns (for 5-day return)

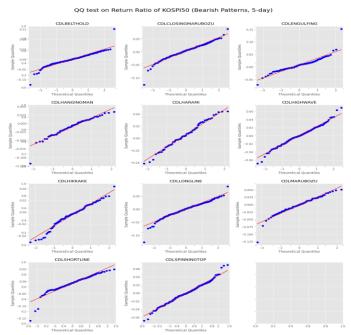


Figure II.3.1.2 Q-Q plot of returns for KOSPI50 bearish patterns (for 5-day return)

Furthermore, some boxplots of data were generated to check if there are significant outliers that violate the t-test assumption. Figure II.3.1.3 and Figure II.3.1.4 show boxplots of HSI and KOSPI50 simple returns for some patterns. Since there are only few outliers which can be ignored and thus it can be assumed that there is no significant outlier in the data.

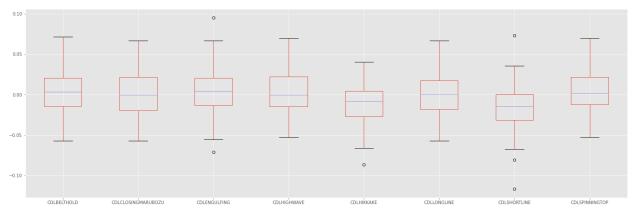


Figure II.3.1.3 Boxplots of simple returns for some patterns observed in Hang Seng Index

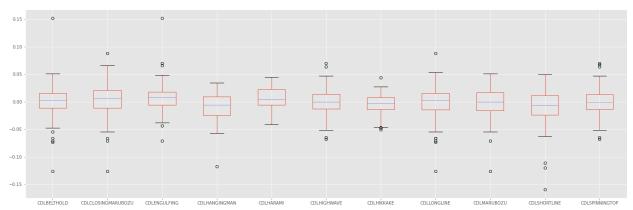


Figure II.3.1.3 Boxplots of simple returns for some patterns observed in KOSPI50

## II.3.2 Hypotheses

The thresholds were defined for each of short and long term prediction which refer to 5-day simple returns and 250-day simple returns respectively. The population mean  $(\mu)$  is hypothesized to be 0.1% for short-term return and 5% for long-term return. Now, the hypotheses for bullish and bearish patterns are:

$$H0: \mu = \underline{R}$$
  $H0: -\mu = \underline{R}$   
 $H1: \mu < \underline{R}$   $H1: -\mu > \underline{R}$   
 $<$  Bullish test  $>$   $<$  Bearish test  $>$ 

#### II.3.3 Test statistic

The test statistic for one sample t Test is denoted as t, which is calculated using the following formula:

$$t_{bull} = \frac{\underline{R} - \mu}{S_{\underline{R}}}$$
 where 
$$t_{bear} = \frac{\underline{R} - (-\mu)}{S_{\underline{R}}}$$
 
$$s_{\underline{R}} = \frac{S}{\sqrt{n}}$$
 where 
$$\mu = \text{Proposed constant for the population mean,}$$
 
$$\underline{R} = \text{Sample mean,}$$
 
$$n = \text{Sample size (i.e., number of observations),}$$
 
$$s = \text{Sample standard deviation,}$$
 
$$s = \text{Sample standard deviation,}$$
 
$$s_{\underline{R}} = \text{Estimated standard error of the mean,}$$
 
$$t_{crit, bear} = -t_{0.05, df}$$

If  $t_{\text{bull}}$  is greater than  $t_{\text{crit,bull}}$  or  $t_{\text{bear}}$  is smaller than  $t_{\text{crit,bear}}$  at 5% level, the null hypothesis is rejected and it can be concluded that there is significant evidence to identify an increasing or decreasing market trend after the pattern is found.

## III. Result

#### III.1. Data Collection

Daily price data of HSI and KOSPI50 for 9 years from the day of data collection (6 Nov 2020 to 7 Nov 2011) were collected from Yahoo finance and finance.com in csv file formats. The range was selected since it was the maximum range of KOSPI 50 data available. Some additional data other than ohlc prices and date such as volume were included.

# III.2. Data Preparation

#### III.2.1. Candlestick Chart

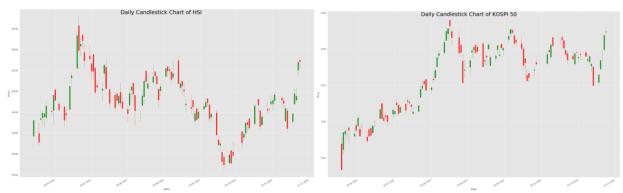


Figure III.2.1.1. Daily Candlestick Chart of HSI(left) and KOSPI50(right)

Figure III.2.1.1. shows the candlestick chart of recent 100 days. KOSPI 50 had rapidly recovered from the deep bottom in early 2020 due to COVID19.

#### III.2.2. Dataframe for Test

From the original data with date and ohcl prices, the simple returns for the days the patterns found constructed to bullish and bearish dataframe separately. Since different parameters in t-test would be applied to short and long term data, daily, weekly and annual dataframes for bullish and bearish of HSI and KOSPI50 had been constructed in total.

|    | CDLBELTHOLD | CDLCLOSINGMARUBOZU | CDLENGULFING | CDLHIGHWAVE | CDLHIKKAKE |
|----|-------------|--------------------|--------------|-------------|------------|
| 5  | NaN         | NaN                | 0.014626     | NaN         | NaN        |
| 12 | NaN         | NaN                | NaN          | 0.007525    | NaN        |
| 13 | NaN         | NaN                | NaN          | 0.001112    | NaN        |
| 15 | NaN         | NaN                | 0.009448     | NaN         | NaN        |
| 16 | -0.020617   | NaN                | NaN          | NaN         | NaN        |

Figure III.2.2.1. First few rows and columns of daily HSI bearish dataframe

#### III.3. Test results

## III.3.1 Bearish patterns found in Hang Seng Index

Figure III.3.1.1 and III.3.1.2 are the results of one sample t-test on short term and long term simple returns of the Hang Seng Index respectively. If t-ratio is smaller than the critical value, we reject the null hypothesis and therefore this means the average simple return for short term trade is less than -0.1% and the average simple return for long term trade is less than -5% at a 5% significance level. As a result, we can see that for short term prediction, **Hikkake** and **Shortline** patterns are the most meaningful patterns. For long term prediction, **Shortline** and **Spinning Top** appeared to be the most powerful predictors.

|                    | n   | mean      | df  | std      | se       | t ratio   | t critical |
|--------------------|-----|-----------|-----|----------|----------|-----------|------------|
| CDLBELTHOLD        | 103 | 0.003478  | 102 | 0.025580 | 0.002520 | 1.776688  | -1.659930  |
| CDLCLOSINGMARUBOZU | 47  | 0.000790  | 46  | 0.028368 | 0.004138 | 0.432642  | -1.678660  |
| CDLENGULFING       | 100 | 0.004087  | 99  | 0.026270 | 0.002627 | 1.936635  | -1.660391  |
| CDLHIGHWAVE        | 40  | 0.002984  | 39  | 0.029388 | 0.004647 | 0.857309  | -1.684875  |
| CDLHIKKAKE         | 93  | -0.010012 | 92  | 0.024509 | 0.002541 | -3.545858 | -1.661585  |
| CDLLONGLINE        | 81  | -0.000533 | 80  | 0.026295 | 0.002922 | 0.159714  | -1.664125  |
| CDLSHORTLINE       | 40  | -0.014959 | 39  | 0.033277 | 0.005262 | -2.653045 | -1.684875  |
| CDLSPINNINGTOP     | 77  | 0.004357  | 76  | 0.026181 | 0.002984 | 1.795336  | -1.665151  |

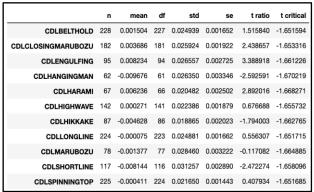
t ratio t critical CDLBELTHOLD 67 -0.046756 66 0.065218 0.007968 0.407172 -1.668271 CDLCLOSINGMARUBOZU 34 -0.044867 33 0.069792 0.011969 CDLENGULFING 84 0.039796 83 0.145922 0.015921 5.639935 -1.663420 CDLHIGHWAVE 26 -0.068436 25 0.071158 0.013955 -1.321062 -1.708141 CDLHIKKAKE 88 0.035519 87 0.133707 0.014253 5.999987 -1.662557 CDLLONGLINE 54 -0.050959 53 0.066336 0.009027 -0.106185 -1.674116 CDLSHORTLINE 29 -0.074793 28 0.069251 0.012860 -1.928013 -1.701131 CDI SPINNINGTOP 56 -0.070470 55 0.079352 0.010604 -1.930387 -1.673034

Figure III.3.1.1 5-day bearish return statistics of HSI

Figure III.3.1.2 250-day bearish return statistics of HSI

## III.3.2. Bearish patterns found in KOSPI 50 Index

Figure III.3.2.1 and Figure III.3.1.2 show the results of t-test on short term and long term simple returns of the KOSPI50 Index respectively. We used the same bearish rejection criterion. For short term prediction, Hanging man, Hikkake and Shortline appeared to be the most powerful indicators whereas none of the **patterns** in long term prediction appeared to be effective.



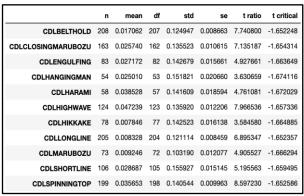


Figure III.3.2.1

Figure III.3.2.2 250-day bearish return statistics of KOSPI 50

5-day bearish return statistics of KOSPI 50

## III.3.3. Bullish patterns found in Hang Seng Index

Figure III.3.3.1 and III.3.3.2 are the results of t-test on short term and long term simple returns of the Hang Seng Index respectively. When testing bullish patterns, we use a bullish rejection criterion that rejects the null hypothesis if t-ratio is greater than the critical value. Rejecting the null hypothesis implies that the average simple return for short term trade is greater than 0.1% and the average simple return for long term trade is greater than 5% at a 5% significance level. As a result, **Hammer** pattern is shown to be the most meaningful short-term predictor. For long-term prediction, **none of the patterns** appeared to be a powerful predictor.

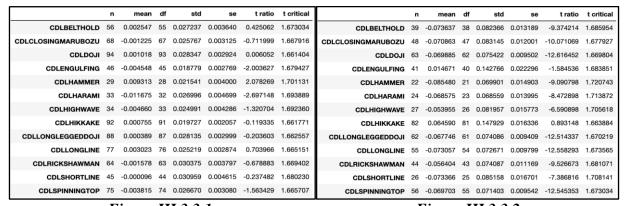


Figure III.3.3.1
5-day bullish return statistics for HSI

Figure III.3.3.2
250-day bullish return statistics for HSI

#### III.3.4. Bullish patterns found in KOSPI 50 Index

Figure III.3.4.1 and Figure III.3.4.2 show the results of t-test on short term and long term simple returns of the KOSPI50 Index respectively. We apply the same bullish rejection criterion. For short term prediction, **Inverted Hammer** appeared to be the most powerful indicator whereas **none of the patterns** in the long term prediction appeared to be effective.

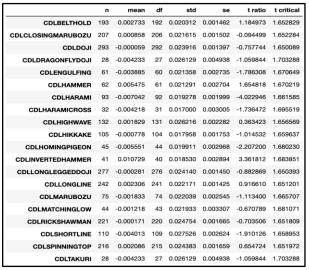


Figure III.3.4.1
5-day bullish return statistics for KOSPI 50

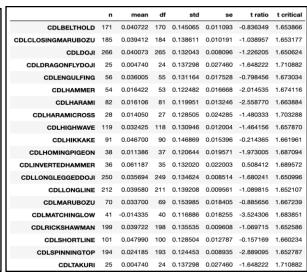


Figure III.3.4.2
250-day bullish return statistics for KOSPI 50

## III.4. Exploration of Findings

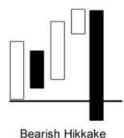
As an investor or trader, statistics and testing would not be their main interests. Introduction to some of the significant bullish/bearish patterns, and their corresponding trading rules will be discussed by the markets (Hong Kong & Korea), and by the periods of time (Short-term & Long-term).

## III.4.1 Hang Seng Index



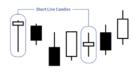
A **Hammer**(Bullish) is a bullish reversal pattern, which signals that a stock is nearing bottom in a downtrend. The body of the candle is short with a longer lower shadow. For the recommended training rule, since the stock is bearish and it is believed the price will go up for a period of time. Therefore, traders should <u>long the asset in a short-term</u> right after a bullish hammer is observed. Appling buy-low-sell-high strategies, traders will be able to earn a profit.

Figure III.4.1.1 Hammer (Bullish)



A **Hikkake**(Bearish) is a complex bar or candle pattern that begins to move in one direction but reverses quickly and is said to establish a forecast for a move in the opposite direction. For the recommended training rule, since the stock is bullish and it is believed the price will go down for a lot. Therefore, traders should short sell the asset in a short-term/long-term right after a bearish Hikkake is observed. Appling buy-low-sell-high strategies, short-sellers can earn a profit by purchasing the asset at a lower price afterward.

**Figure III.4.1.2** Hikkake (Bearish)



A **Short Line**(Bearish) is a short body, and this one-bar occurs when there is only a small difference between the opening and the closing price. For the recommended training rule, it does not provide a straightforward strategy, it indicates a period of consolidation in a stock, but the interpretation will vary based on what other price action has preceded and follows it.

Figure III.4.1.3 Short Line (Bearish)

#### III.4.2. KOSPI 50



An **Inverted Hammer**(Bullish) also forms in a downtrend and represents a likely trend reversal or support. It's identical to the Hammer except for the longer upper shadow. For the recommended training rule, since the stock is bearish and it is believed the price will go up for a period of time. Therefore, traders should <u>long the asset in a short-term</u> right after a bullish inverted hammer is observed. Appling buy-low-sell-high strategies, traders will be able to earn a profit.

Figure III.4.2.1 Inverted Hammer (Bullish)

# **Bibliography**

Farley, A. (2020, September 29). The 5 Most Powerful Candlestick Patterns. INVESTOPEDIA.

https://www.investopedia.com/articles/active-trading/092315/5-most-powerful-candlestick-

patterns.asp

HAYES, A. (2020, March 16). Technical Analysis. INVESTOPEDIA.

https://www.investopedia.com/terms/t/technicalanalysis.asp

HAYES, A. (2020, July 29). Candlestick Definition. INVESTOPEDIA.

https://www.investopedia.com/terms/c/candlestick.asp

mrjbq7. (n.d.). TA-Lib. TA-Lib. https://mrjbq7.github.io/ta-lib/

Rolf, S. (2019, November 16). Mastering and Understanding Candlesticks Patterns. Trading

Strategy. https://www.tradeciety.com/understand-candlesticks-patterns/