NUS INVESTMENT SOCIETY

ACADEMIC RESEARCH REVIEW

Examining Candlestick Patterns through Pattern Recognition

INTRODUCTION

Pattern recognition is a field in Computer Science which attempts to categorize data sets into categories based on common features and to determine the underlying pattern from one category to the next. This field of study is commonly used in speech and handwriting recognition. In this instance, we attempt to harness Pattern Recognition techniques to test commonly held assumptions of candlestick patterns in Technical Analysis.

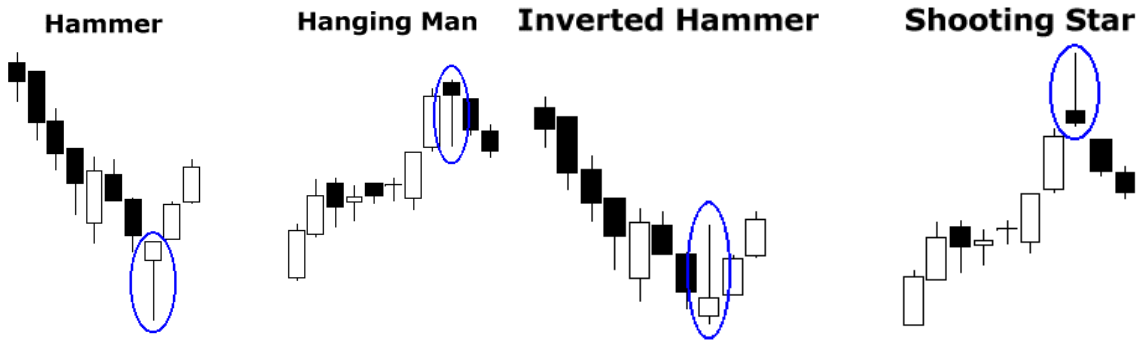

Figure 1. Long Shadow Reversals: Hammers are pre-cursors to bullish markets while Hanging Man and Shooting Stars are pre-cursors to bearish markets

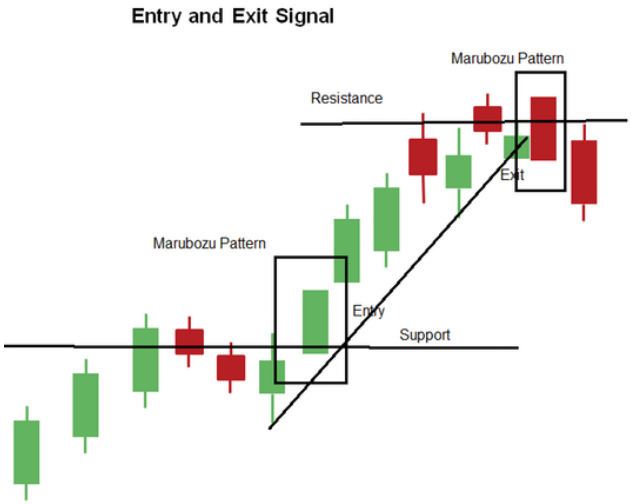

Figure 2. Marabuzu: The highlighted green bar (resp. red bar) indicates that buyers (sellers) controlled the price action from the first to the last trade.

METHODOLOGY

We employ the K-means algorithm in order to test this assumptions. The algorithm goes as follows:

1. Provide the data to be classified (in this case, we use daily Open, High, Low, Close) and the number of categories for the data to be classified in – K.
2. Based on an initial estimate, the algorithm begins with K points and classifies the data based on how close they are to these points.
3. The K points are then recalculated by taking the average of the data points classified into the respective categories.
    a. If the change in the means results in a reclassification of the data points, repeat 2
4. We then estimate the probability of changing from one category to another.

RESULTS

The above algorithm was tested on the Nasdaq 100(QQQQ) index on an arbitrary period with K being set to 6. The choice of the Nasdaq was to eliminate any form of asset or sector specific irregularities by employing a market index.

Figure 3. Plot of the Nasdaq 100 data points used

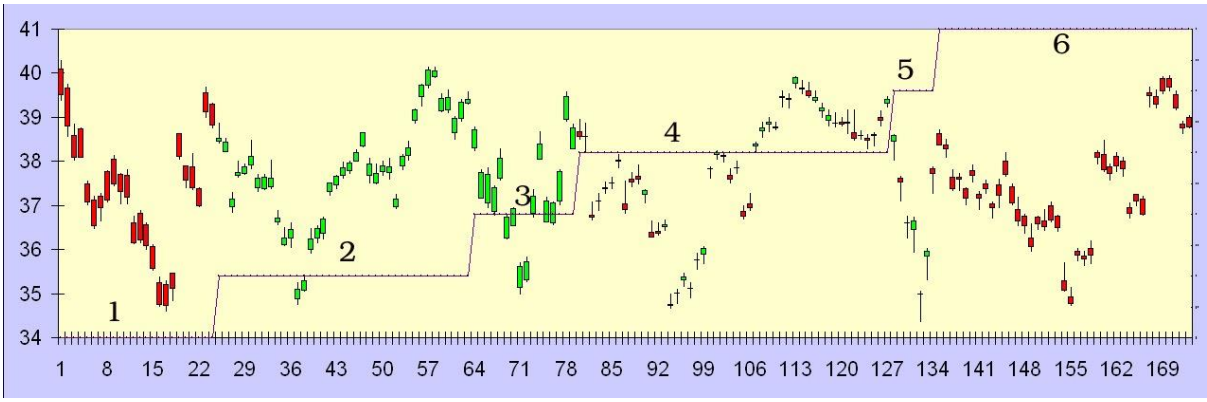

Figure 3. Plot of the Nasdaq 100 after being categorized into 6 groups by the K-means algorithm

At this point, it should be noted that the algorithm does not know the type of candlestick pattern that it is classifying. At no point of time in the algorithm did we specify that the hammers are to be grouped together and the hanging man are to be classified into another group. Instead, we have to manually

label these groups of data with the respective candlestick patterns. It is entirely possible that the algorithm does not classify the data points into a discernible pattern.

Here we note that group 1 resembles a marabuzu when sellers are in control while 3 resembles a marabuzu when buyers are in control. Group 4 resembles dojis while group 5 resembles Long Shadow Reversals. While it is not clear what groups 2 and 6 are, we could postulate that they resemble weaker marabuzu.

		Next day					
		1	2	3	4	5	6
This day	1	20.8	20.8	4.2	37.5	0.0	16.7
	2	13.2	23.7	10.5	34.2	0.0	18.4
	3	12.5	31.2	6.2	18.8	0.0	31.2
	4	8.3	20.8	8.3	22.9	4.2	35.4
	5	57.1	14.3	0.0	14.3	0.0	14.3
	6	7.7	23.1	15.4	28.2	12.8	12.8

Table 1. A probability table indicating the probability of the category the price data will be in the next day given the category it is in today.

CONCLUSION AND EXTENSIONS

From Table 1. we see that there are no significant result that stands out except for 5 → 1 with a probability of 57.1%. Recall that group 5 are Long Shadow Reversals which are used as a pre-cursor of a bearish market. Based on our result, this belief holds as group 1 is the marabuzu where sellers are in control of the market. However, we can see in Figure 3. that the number of occurrences of group 5 type data are extremely infrequent and that the resultant probability is statistically insignificant.

It would appear that candlestick patterns are not applicable to the Nasdaq 100.

From here, we will look into testing the different asset classes and sectors to determine which ones are ideal for using candlestick patterns as a trading strategy. In addition, the current implementation looks at the sequence of group changes in isolation of historical states which is contrary to the techniques used in Technical Analysis. We could extend the current implementation to group the data using a longer time period.

RESEARCH ANALYSTS


Li Dong Yan


Luqman B Lukman


Ooi Gene Yan


Sng Wee Leng Jason