Discovering Pattern Associations in Hang Seng Index Constituent Stocks

Kim Man Lui
Department of Computing, The Hong Kong Polytechnic University
Hung Hom, Hong Kong
E-mail: cskmlui@comp.polyu.edu.hk

Lun Hu (Corresponding author)
Infinite Financial Solutions Limited, Wan Chai, Hong Kong
Tel: 852-6356-4768 E-mail: hulun499@gmail.com

Keith C.C. Chan
Department of Computing, The Hong Kong Polytechnic University
Hung Hom, Hong Kong
E-mail: cskcchan@comp.polyu.edu.hk

Abstract

The problem of finding patterns in financial time series has been tackled by systematic observations of trends, statistical analysis or the use of artificial intelligence techniques in trend analysis. These techniques are more for the discovering of patterns in data rather than the understanding of association relationships between the discovered patterns. As time series patterns often overlap with each other, identifying and discovering association relationships among them can be very challenging. To tackle these problems, we propose here a method to determine if there exists any association relationship between two sequential patterns in a financial time series. The method is based on the use of machine learning techniques and has been tested with data from Hang Seng Index (HSI) constituent stocks. The results reveal that there is statistical evidence of association relationships between some of the stocks whereas there is no evidence for such a relationship between some others. We conclude that the price behavior of these HSI stocks is easier to understand than that of the HSI index.

Keywords: Efficient Market Hypothesis, Hang Seng Index, Price Behavior, Associative relationships, Artificial intelligence

1. Introduction

The efficient market hypothesis (EMH) (Fama, E., 1970) asserts that past price and volume data cannot be used to predict future trend because historical patterns are usually fully considered by people. There are at least three different definitions of the efficient-market hypothesis (EMH) (Clarke, J., Jandik, T. and Mandelker, G., 2001). In this paper, we will use the definition given in (Fama, E., 1970). EMH has been developed based on Mandelbrot's (Mandelbrot, B., 1963) and Fama's early work (Fama, E., 1965). Both of them addressed the price distribution of a market and both suggested that it was not normal but leptokurtic and fat-tailed instead. A recent empirical study by Hamori shows that the monthly data of stock prices in markets in Germany, Japan, the UK and the USA exhibits higher kurtosis values (Hamori, S., 2003). Kaplan suggests, however, that there could be exceptions. He shows that there was a low frequency variability of the exchange rate between the Swiss Franc and the US dollars and argued that the exchange rate may not be a random walk all the time (Kaplan, D.T., 1994).

Based on Kapan's (1994) work, one may question whether EMH and the random walk theory are always true for all financial markets at all time. As different markets have their own characteristics, it is not unreasonably for someone to doubt the validity of these hypotheses and theories. EMH or the random walk is not easily observable in price/volume charts by many investors.

It is well-known that the Futures market is a zero-sum market. However, for stock markets, since it is possible for investors to hold on to particular stocks in order to earn dividend or even to wait for news about merger and acquisition, investing into the stock market is not a zero-sum game (Hadady, R.E., 2000). Many financial derivatives, such as composite index futures, are derived from the values of the stocks that they are made up of. Hence, financial derivatives, such as futures and options, may strongly affect the composite stocks in the stock markets (Chakrabarti R, Huang, W., Jayaraman, N. & Lee, J. 2002). The price behaviours of financial derivatives markets, say future markets, and stock markets can therefore be so different that the EMH may not be equally established all the time, in all markets, and for all stocks.

Technical Analysis (TA) refers to a set of methods that can be used to analyse past price changes in financial data so as to predict future prices (Bulkowski, T.N., 2005). TA has been considered as an indispensable tool by both professional analysts and amateur traders. Even though it has always been criticized to have a lack of sound theoretical foundation (Kirkpatrick, C.D. & Dahlquist, J.R., 2007), it is considered by some to be of fundamental importance in the curriculum of financial studies (Kirkpatrick, C.D. & Dahlquist, J.R., 2007). According to EMH, TA should not be able to reliably speculate what might be happening in the financial markets. Many techniques in TA can theoretically be applied to any financial time series. However, we argue here that the extent to which EMH can be applied to one stock could be different from another. Thus, TA and EMH have been in conflict with each other. In this paper, we present a method to discover association relationship between two price patterns. We tested it with data from the 42 Hang Seng Index constituent stocks and the results show interesting patterns in their price behavior. We believe that it provides a valuable reference in trading.

Price patterns alone may not always provide valuable information for traders. It is sometimes the strong association relationship that one can discover between patterns that can be of interest. Here, we propose a two-stage method that can discover such association relationship. The method consists of a first stage of finding sequential price patterns in an evolutionary manner and a second stage of calculating adjusted residuals (Chan, K.C.C., Andrew, W.K.C. & David C.K.Y., 1994) to determine if some specific patterns have any association relationships with certain time lag. Although there are numerous ways to uncover patterns in financial time series, many of them could be impractical because price patterns themselves often overlap with each other. To uncover such patterns, there is a need for them to be reconstructed from time-series data in an evolutionary manner. To do so, we try to maximize the occurrences of each possible pattern. As long as the data are sufficiently large to allow statistically significant conclusion to be drawn based on statistical analysis, the length of each such pattern is not restricted. The details of how patterns and the association between patterns can be uncovered in time series data are discussed in details here.

The rest of the paper is organized as follows. Section 2 describes how probabilistic patterns can be discovered in time series data using a residual analysis technique. To explain how probabilistic patterns can be discovered, we present a real case using read financial time series data. In Section 3, we present a hybrid method combining evolutionary and a probabilistic approach. The evolutionary algorithm can be used to discover patterns in time series data. The probabilistic approach determines if there is association relationship among patterns. Section 4 shortly reports the implementation and results for analyzing Hong Kong's Hang Seng Index (HSI) and its 42 constituent stocks in 2008. The final section concludes the work.

2. Discovering Probabilistic Patterns

Patterns that do not appear in a time series deterministically can be referred to as probabilistic patterns. These patterns can be discovered in time series data using an effective approach as proposed in (Chan, K.C.C., Andrew, W.K.C. & David C.K.Y., 1994). The method is adopted here to determine association relationships between patterns. The details of it are given as follows.

Given a sequence of events, we would like to determine if some events are inter-dependent on each other. Suppose that a stock can take on a value l, where l can be a qualitative value such as *up*, *no change* or *down*, at time t. Suppose also that the stock takes on the value k at a time lag n from t. In other words, the values l and k are observed at t and at $t+n$, respectively. Let o_{lk} denote the total number of values in the time series that have a k value followed at a time lag n with a l value and let e_{lk} denote the expected number of such values. The adjusted residual can be defined as:

$$d_{lk} = z_{lk}/v_{lk} \tag{1}$$

where v_{lk} is the maximum likelihood estimate of the variance of $z_{lk} = (o_{lk} - e_{lk})/e_{lk}$, and is given by:

$$v_{lk} = (1 - o_{l+}/M^{'})(1 - o_{+k}/M^{'}) \tag{2}$$

where o_{l+} is the total number of data points in the time series that has the characteristic v_i, o_{+k} is the total number of data points in the time series that has the characteristic v_i, and and $M^{'} = \sum_{l,k} o_{lk}$. To be 95% confident that there exists an association relationship, the absolute value of d_{lk} should be more than 1.96 (Haberman, S. J., 1973).

Given the above, we define a Maximum Absolute Adjusted Residual (MAAR) as:

$$\text{MAAR} = \max(|\{d_{lk}\}|) \tag{3}$$

Hence, the MAAR carries the information that if MAAR > 1.96, an association relationship exists between *l* and *k* at time lag *n*, for the given time series on a 95% confidence level.

A real financial data is given for demonstration here. Figure 1 is a closing-price chart. If the closing price compared with its previous one decreases or increases between -0.6% and 0.6%, it is labeled as "L" (level with yesterday). If it decreases by more than -0.6%, it is labeled "D" (down) and if it increases by more than 0.6%, the value is labeled as "U" (up). With such a data transformation, we have a price sequence as shown in Figure 1. Our interest is to determine whether the element at *t* in the series is dependent on that of its previous element (i.e. at *t-n*).

*** Insert Figure 1 Here ***

Using Eqs. (1) and (2), we can construct a contingency table and calculate adjusted residuals. In our example, we use $\text{MAAR}_{n=1}$ to denote that the MAAR calculation is on a symbolic sequence with a one day time lag. Based on the results shown in Table 1, we conclude: (i) there is statistical evidence (as 2.47 > 1.96) that L will come after U, (ii) there is a statistics evidence (as |-2.02| > 1.96) that L will not come after D, (iii) the MAAR is 2.47 which is greater than 1.96 showing that some association relationship exists in the table.

Insert Table 1 Here

3. Association Relationships between Patterns

Many patterns in financial time series can be superimposed or overlapping with each other making them difficult to discover. To deal with such a problem, we have to define atomic elements to construct patterns so that we clearly know which parts of or what elements of a pattern are overlapping. To do so, we use a method consists of the following steps.

Step 1: Given time series data

Step 2: Define a set of symbolic elements

Step 3: Generate a set of atomic patterns.

Step 4: Transform the time series data into a symbolic sequence in terms of atomic patterns

Step 5: Calculate the MAAR of the symbolic sequence

3.1 Symbolic Elements

Given a simple element set *E*={"U", "L", "D"} where U denotes up, L denotes level or unchanged and D denoted Down. In a financial time series, we have to take holidays into consideration because there is no trading activity during holidays and also because (i) price patterns should not be limited to a sequential order. The price patterns can be event-based. For example, some patterns may have frequently happened after or before weekends? (ii) "The performance of stock prices during breaks in trading has received considerable attention in recent years (Fortune, P., 1998)." Therefore, we define E = {"U", "L", "D"} ∪ {"H"}, where "H" denotes holidays and weekends.

For demonstration, let us consider the HSBC bank stock. For example, the time series data in Figure 1 can be converted into a symbolic sequence *s* as {H, U, L, L, D, L, H, L, L, U, L, D, H, D, D, D, U, H, L, D, U, L, L, H, D, D}.

3.2 Atomic Pattern Construction

Atomic patterns are basic building blocks, which are constructed from a simple element set from *E*. Building patterns from such simple elements can be made evolutionary in nature. New patterns can be built from existing simpler patterns until no more new patterns can be built from simpler ones. To guide such iterative algorithm, we have to quantify an optimal solution so that patterns could be ranked against each other. To do so, we consider a subsequence to be a superior individual with a fitness value if it occurs frequently and its length is long. According to such criteria, a fitness function for an evolution algorithm that can find atomic patterns can then be defined as:

$f(c) = f(m, l, o) = 1 \times \ln(o) / (m - l + 1)$ if $o > 0$

$f(c) = f(m, l, o) = 0$ if o=0

where m is the length of s (i.e., the total number of elements of s), l is the length of a subsequence (i.e. the total number of elements of a subsequence) and c is a subsequence of s with l >=2 and o is the occurrences of the subsequence in s.

Given a set E of elements, an evolutionary approach to discovering atomic patterns can be generated using two algorithms. To generate a set of sequences, T, made up of candidates of atomic patterns; and to remove the redundant candidates for a pattern reconstruction map shown in Figure . The details of the two algorithms are given in Figure and 4, respectively.

All elements which have length l=1 are initially considered as seeds of reproduction (i.e. $P \leftarrow E$). The threshold value of minimum fitness for each candidate pattern with length ≥ 2 is an experimentally changeable parameter and the threshold is set at 0.08 in the following example.

*** Insert Figure 2 Here ***

We take the price sequence in section 2 to illustrate how the algorithm in Figure works. Initially, an element is randomly selected as the seed. Searching for “D” in a symbolic sequence s = {H, U, L, L, D, L, H, L, L, U, L, D, H, D, D, D, U, H, L, D, U, L, L, H, D, D}, we have six children with the length=2. Based on the values of the fitness, “LD” and “DD” are selected as parents for the next iteration. We continue the previous procedure and picking children with length =3. “LD” has no child while DD has one, “HDD.” Keep searching for any children until none of them can be found (see Figure). Finally, we have a set of potential atomic patterns, T={“U”, “L”, “D”, “H”, “UL”, “LD”, “DD”, “LL”, “ULL”, “HDD”}.

*** Insert Figure 3 Here***

In our example, we may observe that the candidate “D” occurs in s are overlapped by one of its children “LD”, “DD” and “HDD” and therefore it should be taken out. An algorithm is given in Figure 4. Each candidate is scanned. Some of them will be removed when they can always be covered by others.

Insert Figure 4 Here

Once a set of atomic patterns have been generated, they can then be used to transform a symbolic sequence into an atomic-pattern sequence. For easy reference, we provide a pattern reconstruction map shown in Table 2. Based on the table, s is converted into *Y=”chebcgdeifacehi”* .

Insert Table 2 Here

Transforming a simple symbolic sequence into atomic-pattern sequences has some advantages:

1) The transformation is reversible and it preserves the sequential information in s.

2) The length of Y is smaller than that of s. Therefore, the size of the element set used to construct a sequence Y is larger. We tend to resolve an overlapping problem by having more atomic elements.

3) The original sequence is ordered with respect to time and the atomic-pattern sequence is with respect to an ordered list of occurred patterns.

3.3 Time Lagging and Adjusted Residuals

Any financial time series can be transformed into an atomic-pattern sequence. As in our example, *we have* $Y=\{c,h,e,b,c,g,d,e,i,f,a,c.e,h,i\}$. We will be interested in any association relationship between y_l at t and y_k at $t-n$, where n is a time lag. Based on Eq 3, we can construct a contingency table to discover associative relationships. It should be noted that in our example the equity data has only 22 trading days (i.e. $|s|$=22) and hence, after transformation, the atomic-pattern sequence, which has $|Y|$=15, becomes small. As for implementation, $|s|$ should be large enough so that $|Y|$ is not too small for a statistical analysis.

4. Results and Discussion

We have applied our method to data collected from the Hang Seng Index (HSI), which is a market value-weighted index, as well as its 43 constituent stocks which cover eleven industrial sectors and represent about 65% of the capitalization of the Hong Kong Stock Exchange (Wikipedia, 2009). At the time of implementation of the proposed method, we have available the real-time stock data from Dec 17, 2007 to July 4, 2008 and this was the set of data we used in our experiments to track the changes in price behavior.

Insert Figure 5 Here

In addition, we tested historical stock data from Oct 10, 2007 to Apr 18, 2008. Both results are consistent.

Insert Figure 6 Here

Figure 5 and Figure 6 respectively shows the distribution of maximum absolute adjusted residuals (MAAR) of HSI and each of its constituent stock in different periods. The time lag is varied within the range n={1, 2, 3, 4} and the threshold set as 0.03. The element set is defined as E = {"J", "U", "L", "D", "P"} ∪ {"H"} where "J", short for "Jump up", is used to represent changes of not less than 5%; "U", short for "Up", is used to represent changes between 1% and 5%; "L", short for "Level" is used to represent changes between -1% and 1%; "D", short for "Down" is used to represent changes between -5% and -1%; and "P", short for "deeP down" is used to represent changes of less than -5%. The results are ranked in ascending order. 19 and 26 out of 43 constituent stocks have a MAAR more than 1.96 in Figure 5 and Figure 6 respectively, meaning that there are association relationships between at least two atomic patterns. In contrast, the MAAR of HSI is always low during the same period. It should be noted that the value of absolute adjusted residual is changing with time.

When the markets are efficient and hence historical data cannot be used to forecast the future, the markets will look like random walk (Malkiel, B., 2004). However, the experiment results we have shown in Figures 5 and 6 indicate that the high MAAR is not a rare case as almost half of the stocks have MAAR value more than 1.96.

Since 1960s, empirical evidence on EMH has been mixed, but in general they have not fully supported EMH. In Figures 5 and 6, the fact that not all equities have strong association relationships among patterns indicates that prices can sometimes be much more predictable than at others. HSI is a weighted metric of a number of equities and the regularity found in some stocks will be merged with irregularity. As a result, we hardly discover any association relationship in HSI. It should reflect a situation in which the index traded in futures markets should be more random when we observe data in large time intervals. This does not contradict what many day traders believe in: it is much easier to track the index trend within a day than in-between days. It leads to the question of to what extent the prices move so randomly as not to be predictable.

Many financial advisors support trading the market index (West, D., 2008). Every trader should closely watch market indices which reflect both the state of the market and the sentiment of investors. For example, if a market index is going down due to negative economic growth, we can expect that its constituent stocks also go down. In this case, trading the index seems simpler as we need only to forecast the trend of the market. Furthermore, a major market index often contains broad-based asset equities and represents more than about 60% of capitalization of the market; investors are less risky as they do not put all our eggs in one basket.

Some investment consultants advise an opposite way (Forsythe, G., 2008). Investors should build their own portfolios in order to outperform the market indexes. In this paper, on facts that the high MAAR is not by chance and HSI has no association relationship, we support the investment strategy of building a portfolio rather than selecting composite indexes.

To further validate our findings, a survey over the Internet was conducted in Hong Kong and China from May to July in 2009. We sent emails to 62 people and posted to three investment newsgroup to invite people to visit an online questionnaire website http://www.my3q.com/ where we put our questions there. Among many questions related to their demographical background, the core one would be "will you prefer investing in HSI or picking some of its constituent stocks for investment?"

Insert Table 3 Here

Table 3 shows the results. 67.8% of people would prefer investing in Hang Seng Index rather than selecting and buying some of its constituent stocks. This seems to be in accordance with what has been mentioned, i.e, that index trading is widely supported by many investment consultants. HSI is composed of many blue chip stocks in Hong Kong and has been heavily affected by not only its underlying stocks but also future contracts. Thus there could be stronger irregular volatility and this has been supported by our experimental results of the MAARs of HSI which are mostly zeros. Because of the irregularity of HSI shown in this paper, it should be reasonable to believe that it is easier to predict some constituent stocks instead of their composite index.

5. Conclusions

Artificial intelligence techniques have often been used for stock ranking (Yan, R.J. & Ling, C.X., 2007), price prediction (Xiaowei Lin, Zehong Yang & Song Y., 2009) or markets trend (Sheta, A., 2006 and Rimcharoen, S., Sutivong, D., & Chongstitvatana, P., 2005) and turning-point analysis (Liu, X. Wu, X, Wang, H, & Wang Y., 2008). These applications were developed based on different assumptions about the extent that the stock market follows a random-walk model. In this paper, we are not presenting results investigating into stock price prediction by analyzing patterns in time series data. Instead, since there are few literatures that reports about the price behavior of composite index and its constituent stocks, this paper attempts to advance our understanding of it for investors by proposing to use an evolutionary approach to discover association relationship between patterns in time series data. This approach first uncovers patterns in such data. Once a price sequence in terms of atomic patterns has been found, we can calculate the adjusted residuals to justify whether there are association relationships among the patterns. The statistics analysis shows how some patterns may be associated with the others and whether the results are statistically significant. For future work, we will be investigating into answers to the question "*when might we predict constituent stocks better than composite indexes in a less random market?*"

References

Bulkowski, T.N. (2005). *Encyclopedia of Chart Patterns (2nd Edition)*, Hoboken, N.J.: John Wiley, 2005.

Chakrabarti R, Huang, W., Jayaraman, N. and Lee, J. (2002). "The 'index effect' on stock prices and trading volumes: international evidence," EFMA 2002 London Meetings, available at http://papers.ssrn.com.

Chan, K.C.C., Andrew, W.K.C. and David C.K.Y. (1994). "Learning sequential patterns for probabilistic inductive prediction," *Trans on Systems, Man, and Cybernetics*, Vol 24, No 10, pp 1532-1547.

Clarke, J., Jandik, T. and Mandelker, G. (2001). "The efficient markets hypothesis," *In Expert Financial Planning: Advice from Industry Leaders*, ed. R. Arffa, New York: Wiley & Sons, pp 126-141.

Fama, E. (1965). "The behavior of stock-market prices," *Journal of Business*, vol. 38, No. 1, pp 34-105.

Fama, E. (1970). "Efficient capital markets: a review of theory and empirical work," *Journal of Finance*, vol 25: pp 383–417.

Forsythe, G, "Constructing a diversified stock portfolio," the Summer 2008 edition of On Investing, available at http://www.schwab.com/public/schwab/research_strategies/market_insight/investing_strategies/stocks/constructing_a_diversified_stock_portfolio.html.

Fortune, P. (1998). "Weekends Can Be Rough: Revisiting the Weekend Effect in Stock Prices," Working Papers from Federal Reserve Bank of Boston, No. 98-6.

Haberman, S. J. (1973). "The analysis of residuals in cross-classified tables," *Biometrics*, vol. 29, pp 205-220.

Hadady, R.E. (2000). *Contrary Opinion: Using Sentiment to Profit in the Futures Markets*, Wiley Trading.

Hamori, S. (2003). *An Empirical Investigation of Stock Markets: The CCF Approach*, Kluwer Academic Publishers.

Kaplan, D.T. (1994). "A geometrical statistics for detecting deterministic dynamics," *Time Series Prediction: Forecasting the Future and understanding the past,* Eds. Weigend, A.A. and Gershenfeld, N.A. Addison-Wesley, pp 415-428.

Kirkpatrick, C.D. and Dahlquist, J.R. (2007). *Technical Analysis: the Complete Resource for Financial Market Technicians* Upper Saddle River, N.J.: Financial Times Press, 2007.

Liu, X. Wu, X, Wang, H, and Wang Y. (2008) "Multilayer change-point detection on stock order flows by wavelet transformation," *Proceedings of International Conference on Data Mining Workshops*, ICDMW'08. pp 635-642.

Malkiel, B. (2004). *A Random Walk Down Wall Street (8th ed)*, W.W. Norton & Company.

Mandelbrot, B. (1963), "The variation of certain speculative prices," *Journal of Business*, vol 36, pp 394–419.

Rimcharoen, S, Sutivong, D, Chongstitvatana, P. (2005). "Prediction of the stock exchange of Thailand using adaptive evolution strategies," *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence*, pp 232-236.

Sheta, A. (2006). Software Effort Estimation and Stock Market Prediction Using Takagi-Sugeno Fuzzy Models," *Proceedings of 2006 IEEE International Conference on Fuzzy Systems*, pp 171-178.

West, D. (2008). "Index trading vs. stock day trading," available at http://www.bizcovering.com/International-Business-and-Trade/Index-Trading-Vs-Stock-Day-Trading.391139.

Wikipedia. (2009). "Hang Seng Index," available at http:// en.wikipedia.org/ wiki/Hang_Seng_Index.

Xiaowei Lin, Zehong Yang & Song, Y. (2009). "Short-term stock price prediction based on echo state networks," *International Journal of Expert Systems with Applications*, Vol. 36 , no 3, pp 7313-7317.

Yan, R.J. and Ling, C.X. (2007). "Machine learning for stock selection," *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp 1038–1042.

Table 1. The result of the symbolic price sequence is that $MAAR_{n=1} = 2.47$

Occurrences / *Adjusted Residual*		The Value of Element at *t*		
		U	**L**	**D**
The Value of Element at *t+1*	**U**	0 *-0.94*	4 *2.47*	0 *-1.83*
	L	1 *-0.44*	4 *-0.05*	4 *0.37*
	D	2 *1.24*	1 *-2.02*	4 *1.14*

Table 2. A Pattern Reconstruction Map

Atomic Pattern	Symbol
U	a
L	b
H	c
UL	d
LD	e
DD	f
LL	g
ULL	h
HDD	i

Table 3. Online Survey on Investment Decision

Region	Gender	Age	Investment Experience	Investment Decision#
China	Male	Less than 25	Less than 1 year	H: 1, C: 5
		25 - 50	Less than 1 year	H: 2, C: 3
			1 – 5 years	H: 1, C: 4
	Female	Less than 25	Less than 1 year	H: 3, C: 3
		25 - 50	Less than 1 year	H: 1, C: 1
			1 – 5 years	H: 2, C: 1
	Unknown	Unknown	Unknown	H: 3, C: 6
Hong Kong	Male	Less than 25	Less than 1 year	H: 2
		25 - 50	1 – 5 years	H: 1, C: 3
	Unknown	Unknown	Unknown	H: 2, C: 12
Total	H: 18, C: 38			

#: H denotes Hang Seng Index while C is for some of its constituent stocks.

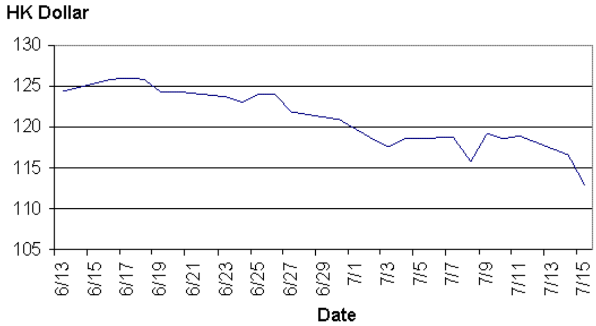

Figure 1. The close price of the Hong Kong and Shanghai Banking Corporation (HSBC; 0005) from June 13, 2008 to July 15, 2008, which can be converted as a finite symbolic price sequence $s=\{U,L,L,D,L,L,L,U,L,D,D,D,D,U,L,D,U,L,L,D,D\}$ for the MAAR calculation. Note that June 13 is only used for its next trading date (i.e. June 16 2008).

1. Input : E and s ;
   Let E be a set of elements;
   Let s be a sequence, $s=\{e_n\}$ where $e_n \in E$;
2. Definition: $|X|$ returns the *cardinality* of a set X ;
3. Initialization: $A \leftarrow E$; $T \leftarrow E$; $C \leftarrow \varnothing$; *threshold* $\leftarrow$ 0.08
4. **do** {
   (1) **for** $i \leftarrow 1$ **to** $|A|$ {
      (a) **for** $j \leftarrow 1$ **to** $|E|$ {
      (b) **if** (a_ie_j **exists** in s and a_ie_j **not exists in** C) **then put** a_ie_j in C
      (c) **if** (e_ja_i **exists** in s and e_ja_i **not exists in** C) **then put** e_ja_i in C
      }
   }
   (2) $A \leftarrow \varnothing$
   (3) **for** $i \leftarrow 1$ **to** $|C|$ {
      (a) **if** $f(c_i) \geq$ *threshold* **then put** c_i in T and in A
      }
   (5) $C \leftarrow \varnothing$
5. } **while** ($A \neq \varnothing$)
6. **return** T

Figure 2. An evolutionary algorithm of generating candidates of atomic patterns

Children of seed D Children of parents LD and DD Children of parent HDD

Parent	Child	Number of Occurrences	Fitness
D	**LD**	**3**	**0.088**
	DL	1	0
	DH	1	0
	DD	**3**	**0.088**
	DU	2	0.055
	HD	2	0.055

Parent	Child	Number of Occurrences	Fitness
LD	LLD	1	0
	LDL	1	0
	ULD	1	0
	LDH	1	0
	HLD	1	0
	LDU	1	0
DD	**HDD**	**2**	**0.92**
	DDD	1	0
	DDU	1	0

Parent	Child	Number of Occurrences	Fitness
HDD	DHDD	1	0
	HDDD	1	0
	LHDD	1	0

Figure 3. Finding Atomic Patterns in an evolutionary manner

1. Input : T and s ;
2. Definition: Let P be a set of atomic patterns;
   Let l_{min} be the minimum length for selecting the next atomic pattern;
   $|X|$ returns the *cardinality* of a set X;
   |s| returns the length of the sequence s;
   $|t_j|$ returns the length of a pattern t_j;
   substring(s, P, L) is a function of a sequence s that returns a subsequence of s which starts at the position P of s and ends at the position (P+L) of s.
3. Initialization: $l_{min} \leftarrow 0$; $P \leftarrow \varnothing$; $M \leftarrow \varnothing$
4. **for** $i \leftarrow 1$ **to** $|s|$ {
   (1) **for** $j \leftarrow 1$ **to** $|T|$ {
   (a) **if** t_j = **substring**(s , i , $|t_j|$) **then put** t_j in M
   }
   (2) $m \leftarrow$ the longest element in M
   (3) **if** the length of m > l_{min} **then**
   (a) $l_{min} \leftarrow$ the length of m
   (b) **if** (m **not exists in** P) **then put** m in the P
   (4) $l_{min} \leftarrow l_{min}$ - 1
   (5) $M \leftarrow \varnothing$
   }
5. **return** P

Figure 4. An algorithm of filtering candidates of atomic pattern

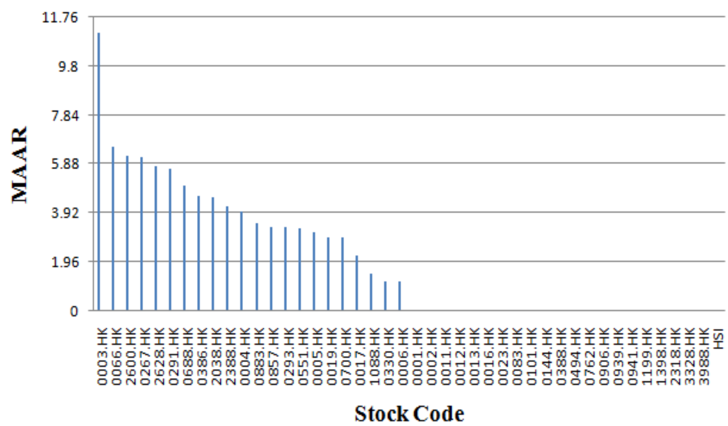

Figure 5. MAAR$_{n=1,2,3,4}$ Distribution for HSI and its constituent stocks in descending order (Dec 17, 2007 to July 4, 2008)

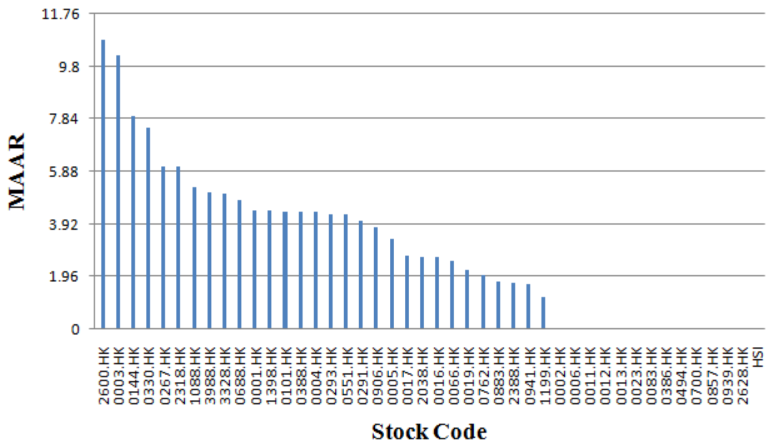

Figure 6. MAAR$_{n=1,2,3,4}$ Distribution for HSI and its constituent stocks in descending order (Oct 10, 2007 to Apr 18, 2008)