Proceedings of the International MultiConference of Engineers and Computer Scientists 2016 Vol I,
IMECS 2016, March 16 - 18, 2016, Hong Kong

Stock Selection and Trading Based on Cluster Analysis of Trend and Momentum Indicators

Ratchata Peachavanish

Abstract—**Stock selection in active portfolio management is a critical task if one wishes to outperform the market. Many research efforts have been made to profitably pick stocks and create trading rules using both fundamental analysis and technical analysis. This paper proposes a method using cluster analysis to identify a group of stocks that has the best trend and momentum characteristics at a given time, and therefore are most likely to outperform the market during a short time period. Experimental results show that using the proposed method to select stocks from the Thai stock market, and trade them using equal-weight monthly portfolio rebalancing, market outperformance can be obtained in the long run.**

Index Terms— **technical analysis, stock market, clustering, trend, momentum**

I. Introduction

STOCK market analysis is an important and popular application in data mining where the primary goal is to create profitable algorithmic trading systems that can be executed without intermediate human judgements. A trading system needs to make frequent, periodic decisions on which stocks to buy or sell and when to trade them. This active trading approach differs from the typical buy-and-hold strategy where investors buy a selected group of stocks (or every stock on the market, e.g., through an Exchange-Traded Fund) and hold on to it for a relatively long period of time without any trading. Actively-managed mutual funds use active trading approach and are judged by their ability to outperform the market index; although most are not able to do so over the long run [1].

In active trading, there are generally two schools of thoughts in stock analysis, *fundamental analysis* and *technical analysis*. Fundamental analysis examines the actual business entity represented by the stock under consideration. The nature of the business, its finance, and its competitiveness are some of the important criteria needed to be analyzed critically. The intrinsic value of the business is determined and compared to the market value of the business as reflected by its current stock price; only then the decision is made whether to buy or sell the stock based on the expected future price reflective of the business prospect. On the other hand, technical analysis only considers the market price series of the stock. It assumes that the market price is fully discounted by the business fundamentals and is govern solely by the demand and supply from market participants [2]. An important assumption in technical analysis is that *price moves in trend and has momentum*; and as a consequence, it tends to continue on its current direction until reversal.

Stock traders may choose to combine both fundamental analysis and technical analysis in making trading decisions. Furthermore, modern stock markets are now participated by automated computerized systems that make trading decisions and execute them in real time without human intervention. These computerized traders use sophisticated algorithms that apply both fundamental analysis and technical analysis and are now account for the majority of trading volume in stock markets in developed countries [3].

This paper proposes a method for stock selection and portfolio rebalancing using cluster analysis on trend and momentum technical indicators. An experiment was performed on 5-year historical price data of stocks listed on the *Stock Exchange of Thailand (SET)*. Section II reviews related works and methods on algorithmic trading, technical analysis, and cluster analysis. Section III describes the proposed method. Section IV and V discusses experimental results and conclusion, respectively.

II. Related Works and Methods

A. Algorithmic Trading

There is a large body of works related to algorithmic stock trading. Generally, they can be categorized by their analytical approaches – fundamental analysis or technical analysis. Supervised machine learning techniques such as support vector machines and neural networks, as well as evolutionary computations such as genetic algorithms, were widely applied to data from stock markets around the world with different degrees of success. Business fundamental data were used in [4] and [5] for stock selection with results that outperform their respective indices. Data derived from technical analysis were used in [6], [7], [8], and [9] to generate trading rules. A comprehensive review of stock algorithmic trading methods [10] found that most perform effectively in downtrend markets and poorly in uptrend markets, and transaction cost is a significant factor.

In addition, other interesting approaches to stock market prediction also appear in the literature. For example, social network data from *Twitter* were used to predict daily change of the Dow Jones Industrial Average index [11].

Despite the rich literature on the topic, it was difficult to compare the different methods due to many factors. Developed markets like those represented by the American S&P500 index are very different from emerging markets like

Manuscript received December 8, 2015. Ratchata Peachavanish is with the Department of Computer Science, Thammasat University, Pathum Thani 12121, Thailand. E-mail: rp@cs.tu.ac.th.

Thailand. For example, regulations and enforcement against insider trading in the Thai market are different from those in developed market [12], [13], leading to abnormal price characteristics (e.g., price goes up or down in extreme with no apparent reason). And although no hard evidence exists, stock price manipulations are believed to be the norm. Additionally, foreign fund-flows have disproportionate effect on market return and volatility in the relatively small Thai market [14], making fund flow information critical in decision making.

Furthermore, objectives of different research studies vary widely. While some researchers aimed to select a set of stocks from a large, diverse pool, others concentrated on a single price series, such as an index, and try to find profitable trading rules or predict future price of that specific price series [15]. Accordingly, a more straightforward way to judge the performance of an algorithm is to measure it against the market index return, which is normally how mutual funds are judged. For example, a fund that buys all stocks in the market on an equal-weight basis instead of using market-capitalization weight often outperforms the market [16] and is advertised as such. More advanced *Smart Beta* mutual funds select stocks using a set of quantitative criteria, such as dividend yield and volatility, without fundamentally analyzing businesses behind the stocks [17].

B. Technical Analysis

There are numerous aspects in technical analysis of stock price series. A technical analyst may try to predict future price direction based on graphical patterns and shapes of price charts. Also, trading volume may be considered to gauge the strength of price movement. Lacking formal taxonomy, technical analysis methods are grouped by [10] into the following categories: sentiment, flow-of-funds, raw data, trend, momentum, volume, cycle, and volatility.

A common and popular technical trading strategy is *trend following* [18] where instead of trying to predict the future, trading decisions are made *reactively* to the price movement. In this strategy, trend direction and strength of a price movement are computed from historical data, and profits are made by trading in the same direction of the trend (i.e., buy on uptrend, sell or short on downtrend). Since the assumption is that a trend tends to continue on the same direction, the strategy is *"buy high and sell higher"*. This is the approach taken by this paper where two types of technical indicators were used: *trend* and *momentum*.

The moving average is a form of trend indicator when used in conjunction with price. It smooths a price series and gives direction of a trend, albeit with lag. Moving averages can be computed in many ways, each with different characteristics. This paper uses the *Exponential Moving Averages (EMA)* defined in (1), which is a type of moving average indicator that gives more weight to recent prices [2]. The EMA value for day *i* incorporates the day's price as well as the EMA value of the previous day. The variable *Period* in the equation is how far back in the past the data should be used for computation. A relatively large *Period* value signifies a relatively long-term indicator.

$$K = 2 / (Period + 1)$$

$$EMA_i = EMA_{i-1} + K * (Price_i - EMA_{i-1}) \quad (1)$$

Typically, if the current day price is above its moving average, the stock is considered to be in an uptrend; otherwise it is in a downtrend. In cases where multiple moving averages with different *Period* values are used, long-term trend and short-term trend may conflict. The situation occurs when the current price is simultaneously above its long-term moving average and below its short-term moving average (Figure 1). This often occurs when an uptrend stock is temporarily experiencing retracement or correction (i.e., temporary price decrease).

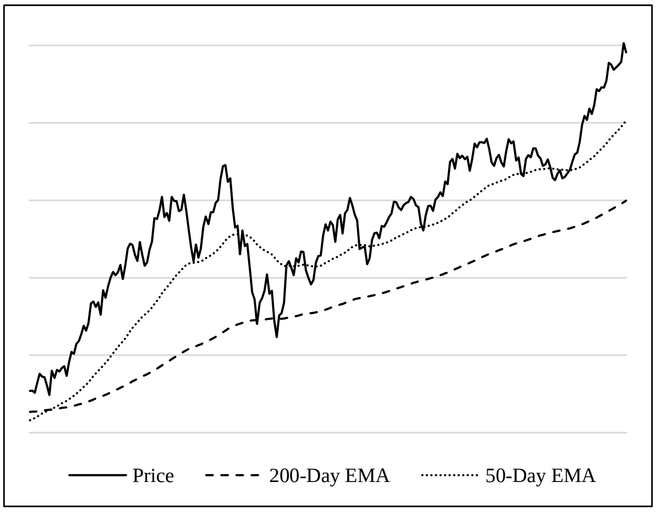

Fig. 1. Long-term (200-day) and short-term (50-day) exponential moving averages overlaid on price series. Stock on a long-term uptrend experienced temporary retracement when its price fell below a short-term moving average.

For momentum, the Rate of Change (ROC) indicator simply measures the price change, in percentage, between the price n days ago and the current price (2).

$$((Price_{today} - Price_{n\ days\ ago}) / Price_{n\ days\ ago}) * 100 \quad (2)$$

Positive and increasing ROC signifies acceleration of price increase. Negative and decreasing ROC signifies acceleration of price decrease. Generally, ROC measures the strength of the trend.

C. Cluster Analysis

Cluster analysis is an unsupervised method for dividing data into groups that are meaningful. Often, it is used as a starting point for further computation as is the case for this paper. A simple partitional clustering simply divides data objects into non-overlapping subsets. A prominent algorithm for partitional clustering is *K-means* where data are grouped into a predetermined number of clusters specified by user. The basic K-means algorithm for clustering into K groups is as follows [19]:

Choose K data points as initial centroids.
Repeat
Form K clusters by assigning each data point to its closest centroid.
Re-compute the centroid of each cluster.
Until Centroids do not change.

K-means algorithm has been used as part of a process for stock trend prediction. For example, [20] used K-Means to aggregate similar technical charts in order to construct trading rules.

III. Method

A. Data Preparation

As of September 2015, there were over 600 listed companies in the Thai stock market with total market capitalization of over THB 13 trillion. The main market index, the *SET Index*, is calculated from the stock price of all listed companies using market-capitalization weight. The exchange also provides the SET50 and SET100 indices, respectively comprising top 50 and 100 stocks chosen based on size and liquidity [21]. Constituents of SET50 and SET100 are regularly revised every six months.

The universe of 162 stocks in this analysis comprises all stocks that have historically ever been included as members of SET100 constituents from January 2011 to September 2015. All price data have been adjusted for any changes in par values.

In order to compare price trend and momentum of different stocks, the following four technical indicators were calculated. For each day, two trend indicators were computed, one for the long term trend and the other for the short term trend. Each indicator is the percentage difference between the daily closing price and its exponential moving average (3).

$$((P - EMA) / EMA) * 100 \qquad (3)$$

The 200-day exponential moving average was used for the long term indicator and the 50-day exponential moving average was used for the short term indicator.

For momentum indicators, 125-day and 20-day price Rates of Change were used for the long-term and the short-term momentum indicators, respectively. The time periods chosen for all indicators were based on [22].

B. Portfolio

The exchange requires a minimum lot size of 100 shares per trade. To make simulation realistic, a large starting amount of cash is required to minimize slippage, which is the difference in quantity and price between the actual trade and what the algorithm specifies. For example, if a stock is priced at THB 400 per share, a buyer must buy multiples of 100 shares or multiples of THB 40,000, at a minimum. For this experiment, the starting amount of cash was THB 100 million. In addition, for each and every transaction (buying or selling), a 0.1578% brokerage commission fee, which is the normal rate for retail investors, was applied.

On the first day, the portfolio consisted of all cash, with no stocks on hand. At the end of the day, when the closing price of all stocks were available, stocks were ranked using the method described in the next section. Using all available cash, the top 10 stocks in the ranking were bought on the following day at the opening price on an equal-weight basis. The portfolio was then rebalanced once every 20 trading sessions (approximately one month) by re-computing the ranking and making necessary buy/sell orders to maintain top-10 stocks allocation, equally weighted.

C. Stock Ranking

For this experiment, only stocks that were members of the SET100 constituents on the day of the ranking were considered. By definition, SET100 comprises 100 largest and most liquid stocks in the market and therefore are most likely to be investment-grade and least likely to be maliciously manipulated. The task was to use trend and momentum indicators to select a subset of these stocks that are most likely to outperform the market in the near term. Clustering was performed on five attributes: long-term trend indicator, short-term trend indicator, long-term momentum indicator, short-term momentum indicator, and current market capitalization of the stock. The inclusion of the market capitalization as an attribute is based on a well-known observation that large stocks tend to move slower than small stocks and thus should be useful in the clustering process. For each attribute, data values were normalized.

To rank stocks for day *i*, technical indicator values calculated on day *x-20* (20 trading sessions ago) were used in the clustering process. K-means clustering was performed with 10 target clusters, corresponding to an average of 10 stocks per cluster. Initial centroids were the 10 stocks that were placed equally far apart when ranked by their long-term trend indicator.

The resulting clusters of stocks formed 10 portfolios, each of which is calculated for profit/loss as if it was held from day *x-20* to day *x*. The centroid of the most profitable (or least lost) cluster is then chosen as the best centroid. Finally, the stock ranking is the ascending order of stocks based on their distances to the best centroid (closest stock ranks highest) calculated on day *x-20*. The rationale behind this ranking is that, for a given market condition, the trend and momentum characteristics of the best-performing group of stocks will continue in the same direction.

IV. Results

The overall result spanning January 2011 to September 2015 is shown in Figure 2. This is the result in which the portfolio was fully invested in stocks at all times with no cash holding. Specifically, the SET index generated 29.0% returns over the period, while the proposed method generated 104.0% return over the same period. It should be noted that the compounding effect of reinvested profits greatly enhanced the return over the long run. The total commission fee paid was THB 26.5 million and was subtracted from the portfolio at the time of each transaction.

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

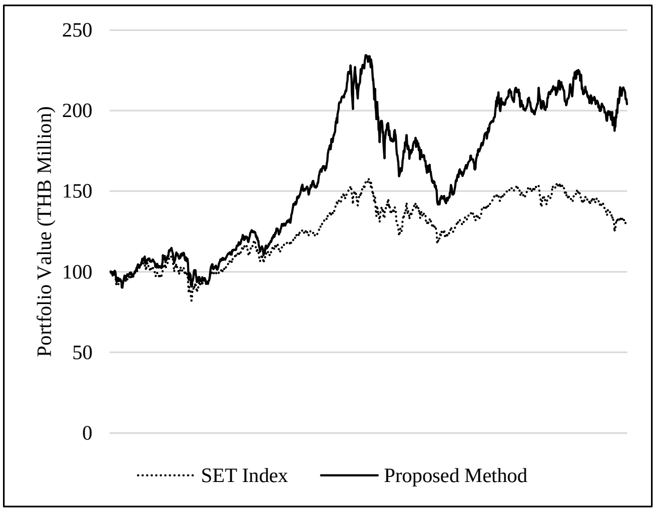

Fig. 2. Portfolio value during the period of 19 consecutive quarters. SET index return is provided for comparison. The return is greatly enhanced due to compounding effect of profit reinvestment.

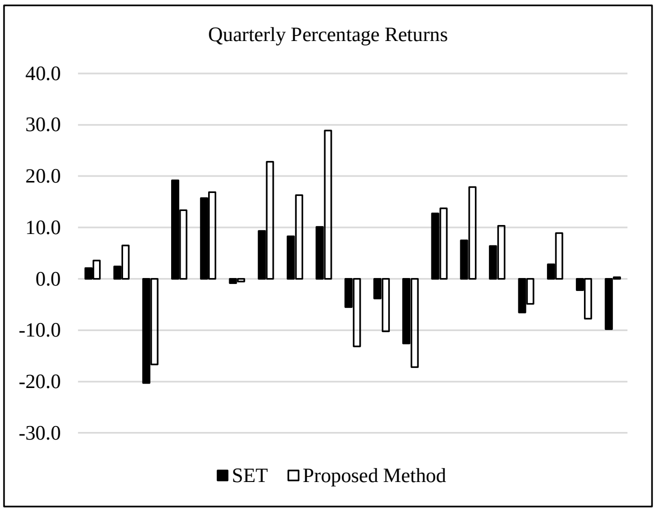

Fig. 3. Quarterly returns using the proposed method compared to the SET index. During uptrends, the proposed method strongly outperformed the index. The method underperformed during high-volume sell offs.

To remove long-term compounding effect, the results are examined on a quarterly basis and is shown in Table 1 and Figure 3. Average quarterly return from the SET index was 1.8% versus 4.7% obtained from the proposed method.

The proposed method underperformed the index during the periods of high-volume sell-offs. Events corresponding to the major sell-offs were the year 2011 Thailand's country-wide flooding and the year 2013 political instability that led to the coup d'état. These results are consistent with trend following strategy as the stocks that have greatly outperformed in the past suddenly made reversals and reverted back to the mean when exceptional events occurred. And since the nature of technical indicators used in the proposed method lag behind prices, it took some time for the selection process to recover and outperform the market again.

TABLE I
QUARTERLY PERCENTAGE RETURNS

Quarter	SET	Proposed Method	Outperformance
1st 2011	2.1	3.6	1.5
2nd 2011	2.4	6.5	4.1
3rd 2011	-20.3	-16.7	3.6
4th 2011	19.2	13.4	-5.8
1st 2012	15.7	16.9	1.2
2nd 2012	-0.9	-0.5	0.3
3rd 2012	9.3	22.8	13.5
4th 2012	8.3	16.3	8.0
1st 2013	10.1	28.9	18.8
2nd 2013	-5.5	-13.2	-7.6
3rd 2013	-3.8	-10.2	-6.4
4th 2013	-12.6	-17.2	-4.6
1st 2014	12.7	13.7	1.0
2nd 2014	7.5	17.9	10.3
3rd 2014	6.4	10.3	3.9
4th 2014	-6.6	-4.9	1.7
1st 2015	2.9	8.9	6.0
2nd 2015	-2.2	-7.8	-5.5
3rd 2015	-9.8	0.3	10.1
Average	1.8	4.7	2.8

Figure 4 shows how performance degraded when number of stocks in the portfolio increased. There were 100 stocks to choose from the SET100 constituents. If the portfolio consisted of half of these stocks, performance degraded to near that of the index return. However, if less than 5 stocks were chosen, the return greatly increased, along with risks specific to companies and sectors of the holdings. Nevertheless, the increase in return when fewer stocks were held demonstrated that the proposed method is effective in picking out market-outperforming stocks.

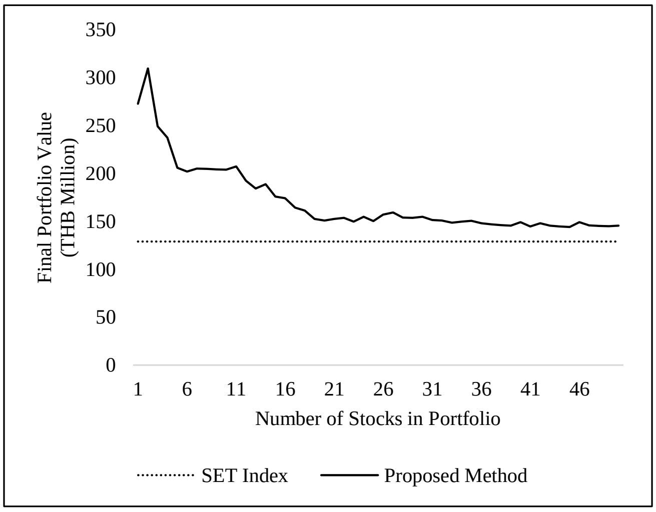

Fig. 4. Final portfolio value after 19 consecutive quarters for different numbers of stocks held in portfolio. The more diverse the portfolio, the less the return. Holding less than 5 stocks greatly increase the return, but with less diversification and more risk.

V. CONCLUSION AND FUTURE WORK

This paper contributes a method that uses a combination of long-term and short-term, trend and momentum technical indicators to identify stocks that are most likely to outperform the market index. The best combination of these indicators are determined by way of cluster analysis. The results show that the proposed method can outperform the market in the long run.

To improve, additional works need to be done to identify reversals of price trend. In addition, the trend following approach utilized in this paper may not work well in a bear market, in which the Thai market has not recently experienced. A combination of short sales and derivative instruments may be needed to make profit in such an environment.

REFERENCES

[1] C. Mateepithaktham, "Equity mutual fund fees & performance," SEC Working Papers Forum, The Securities and Exchange Commission, Thailand, June 2015.

[2] J. Murphy, *Technical analysis of the financial markets*, New York Institute of Finance, p 23, 1999.

[3] A. Haldane, "Patience and finance," Oxford China Business Forum, Beijing, September 2010.

[4] H. Yu, R. Chen, and G. Zhang, "A SVM stock selection model within PCA," 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014.

[5] C. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Applied Soft Computing*, 12 (2012), pp. 807-818.

[6] A. Esfahanipour and S. Mousavi, "A genetic programming model to generate risk-adjusted technical trading rules in stock markets," *Expert Systems with Applications*, 38 (2011), pp. 8438-8445.

[7] T. Chavarnakul and D. Enke, "A hybrid stock trading system for intelligent technical analysis-based equivolume charting," *Neurocomputing*, 72 (2009), pp. 3517-3528.

[8] J. Larsen, "Predicting Stock Prices Using Tehnical Analysis and Machine Learning," Thesis, Norwegian University of Science and Technology, June 2010.

[9] A. Sundén, "Trading based on classification and regression trees," Thesis, Kungliga Tekniska högskolan Institutionen för matematik, 2010.

[10] Y. Hu, K. Liu, X. Zhang, L. Su, E.W.T. Ngai, M. Liu, "Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review," *Applied Soft Computing*, 36 (2015), pp. 534-551.

[11] J. Bollen, H. Mao, X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, 2 (2011), pp. 1-8.

[12] W. Laoniramai, "Insider Trading Behavior and News Announcement: Evidence from the Stock Exchange of Thailand," CMRI Working Paper 3/2013, The Stock Exchange of Thailand.

[13] K. Kwandham, "Insider trade filing and earnings announcement: evidence from the stock exchange of Thailand," *International Journal of Technical Research and Applications*, Special Issue 16 (Jan-Feb 2015), pp. 69-79.

[14] C. Chotivetthamrong, "Stock market fund flows and return volatility," Ph.D. Dissertation, National Institute of Development Administration, Thailand, 2014.

[15] E. Guresen, G. Kayakutlu, T.U. Daim, "Using artifical neural network models in stock market index prediction," *Expert Systems with Applications*, 38 (2011), pp. 10389-10397.

[16] Y. Plyakha, R. Uppal, and G. Vilkov, "Equal or Value Weighting? Implications for Asset-Pricing Tests," World Finance Conference, Venice 2014.

[17] N. Amenc, F. Goltz, "Smart Beta 2.0," *The Journal of Index Investing*, Winter 2013, Vol. 4, No. 3: pp 15-23.

[18] M. Covel, *Trend following: how great traders make millions in up or downmarkets*, Financial Times Prentice Hall Books, 1st ed., New Jersey, 2004, pp. 336.

[19] P. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*, 1st ed. 2005, Pearson, p 497.

[20] K. Wu, Y. Wu, H. Lee, "Stock trend prediction by using K-means and AprioriAll algorithm for sequential chart pattern mining,", *Journal of Information Science and Engineering*, 30 (2014), pp. 653-667.

[21] *SET50 and SET100 Indices Rule*, Avalable: https://www.set.or.th /en/products/index/files/2013-01-SET50-100-IndexRule-EN.pdf

[22] StockCharts Technical Rank, Available: http://stockcharts.com /school/doku.php?id=chart_school:technical_indicators:sctr