Genetic programming optimization for a sentiment feedback strength based trading strategy

Steve Y. Yang[a,*], Sheung Yin Kevin Mo[a], Anqi Liu[a], Andrei A. Kirilenko[b]

[a] *Stevens Institute of Technology, Financial Engineering Program, 1 Castle Point on Hudson, Hoboken, NJ 03070*
[b] *MIT Sloan School of Management, 50 Memorial Drive, Cambridge, Massachusetts 02142*

Abstract

This study is motivated by the empirical findings that news and social media Twitter messages (tweets) exhibit persistent predictive power on financial market movement. Based on the evidence that tweets are faster than news in revealing new market information, whereas news is regarded broadly a more reliable source of information than tweets, we propose a superior trading strategy based on the sentiment feedback strength between the news and tweets using generic programming optimization method. The key intuition behind this feedback strength based approach is that the joint momentum of the two sentiment series leads to significant market signals, which can be exploited to generate superior trading profits. With the trade-off between information speed and its reliability, this study aims to develop an optimal trading strategy using investors' sentiment feedback strength with the objective to maximize risk adjusted return measured by the Sterling ratio. We find that the sentiment feedback based strategies yield superior market returns with low maximum drawdown over the period from 2012 to 2015. In comparison, the strategies based on the sentiment feedback indicator generate over 14.7% Sterling ratio compared with 10.4% and 13.6% from the technical indicator-based strategies and the basic buy-and-hold strategy respectively. After considering transaction costs, the

*Corresponding author
Email address: steve.yang@stevens.edu (Steve Y. Yang)

sentiment indicator based strategy outperforms the technical indicator based strategy consistently. Backtesting shows that the advantage is statistically significant. The result suggests that the sentiment feedback indicator provides support in controlling loss with lower maximum drawdown.

Keywords: News sentiment; Tweet sentiment; financial market; feedback; genetic programming

1. Introduction

The fundamental role of investor sentiment on market anomalies has been well documented in the field of behavioral finance [1, 2]. Studies have shown that sentiment is linked to investor's cognitive and psychological traits and has impact towards financial market movement [1]. With the increasing digitization of textual information, news and social media have become major resources that investors use to gather information on important financial events and to make their corresponding investment decisions. This changing landscape of the way information is delivered has prompted the growing influence of news and social media among multiple stakeholders. For instance, major media publishers such as the Wall Street Journal and the Associated Press use Twitter to disseminate headlines of breaking and regular news to their subscribers. Financial data vendors including Bloomberg and Thomson Reuters incorporate feeds from Twitter and various news sources to meet the demand of clients who want to receive and analyze the most up-to-date and reliable information. On the receiving end, there are numerous claims that high frequency traders and hedge funds are actively monitoring Twitter and news feeds for trading signals. Moreover, an increasing linkage between social media and financial markets has been observed where a number of individual tweets between 2011 and 2013 were found to trigger abrupt market movements [1].

[1] The April 23, 2013 flash crash triggered by the Associated Press Hoax incident is a good example that demonstrates the direct relevance of social media in the financial market. At 1:07pm, the Associated Press (AP) Twitter account tweeted a malicious message regarding

This study is motivated by three main areas of research findings. First, the mechanism of how sentiment affects financial market movements has been studied in the form of theoretical framework and empirical evidence. Barberis et al. [2] initially developed a theory of investor sentiment to illustrate the effect of investor overreaction and underreaction to public information on generating post-earnings announcement drift, momentum and long-term reversals. Daniel et al.[3, 4] further enriched the theory with the psychological premise that investors with private information are overconfident about its precision. On the empirical front, a number of studies found quantitative measures of investor sentiment significant in explaining asset price and volatility movements. Chopra et al. [5] showed that prior losing portfolios significantly outperform prior winning portfolio by 5-10% annually for 5 years, validating the overreaction effect, while Porta et al. [6] displayed evidence that the correction of the extreme investor sentiment tends to revert during earnings announcements when investors realize their initial beliefs were too extreme. Shleifer [7] pointed out that investor sentiment influences prices and the inefficiency of the financial markets are evident across theoretical and empirical literature. In more recent studies, Tetlock [8] argued that negative expressions in news stories have stronger correlation to stock market than positive ones. According to the finding, Tetlock et al. [9] quantified investor sentiment as the fraction of negative words in news stories, and justified the predictability of investor sentiment to individual company's stock price movements with news from Dow Jones News Service (DJNS) and Wall Street Journal (WSJ). In a similar study, Engelberg et al. [10] indicated higher abnormal returns of short sells based on news events in the Dow Jones archive. Baker et al. [11] showed that investor sentiment for major stock markets has

an attack to the White House that President Obama was injured. The message was found to be a hoax with rapid spread on the social media platform. Subsequently, it exert significant downward pressure on the U.S. stock market, which suffered a large intraday decline of more than 2%. Within minutes, the market quickly rebounded to its original level after it was determined that the AP account was hacked.

predictive power of the cross-sectional returns within markets, and Brown and Cliff [12] demonstrated that investor sentiment predicts market returns with its explanatory power on the deviations of stock prices from intrinsic value. García [13] tracked the New York Times financial news columns from 1905 to demonstrate that news content is more robust in predicting stock returns in recessions. Kurov [14] further illustrated the impact of investor sentiment on monetary policy decisions and the stock market. These studies are instrumental in demonstrating the existence of investor sentiment along with its impact on the financial markets.

Furthermore, the second area of literature focused on the empirical observations that media is an important factor of influencing investor sentiment [15], and news and tweets sentiments exhibit persistent predictive power on financial market movement. For tweets sentiment, Bollen et al. [16] showed that tweet messages have shown an accuracy 87.6% in predicting changes in DJIA with a reduction of prediction error. Zhang et al. [17] further showed that the emotional outbursts of tweet activities can predict the next day movement in the financial market. Our previous study constructed a financial community in the Twitter universe where its constituents' interests are aligned with the financial market, and we found that their tweet sentiment has significant correlation with market returns and volatility [18]. On the other hand, a number of empirical studies have demonstrated the significance of news sentiment towards the financial market. Li et al. [19] quantified the media influence on the market and concluded that news sentiment has a notable impact on the emotions and decision-making of investors. Piškorec et al. [20] developed a measure of collective behaviors based on financial news and showed that a news-based index can be used as a volatility indicator. The authors further illustrated that the cohesiveness in financial news has high correlation with market volatility [20]. In addition, corporate news events related to earnings announcements exhibit clustering behavior and trigger significant short-term price changes [21]. Smales [22, 23] illustrated that the empirical sentiment series can explain market returns and volatility. In a former study, we presented evidence that there exists a feedback mechanism between news sentiment and market returns among the major U.S. financial market indices, namely S&P 500, NASDAQ, and Dow Jones Industrial Average [24].

As a natural extension of these empirical findings, there has been a growing number of academic studies that showcase the potential of using sentiments for developing and implementing trading strategies with advanced statistical methods. Dempster and Jones [25] developed a real-time quantitative trading system based on six technical indicators and it generates positive returns with statistical significance. Tetlock [8] developed a trading strategy based on the content of each firm's news stories during the prior trading day, and concluded that the negative fraction of the media content is a significant factor in earning substantial risk-adjusted returns. On a related study, Khadjeh Nassirtoussi et al.[26] applied a multi-layer dimension reduction algorithm on breaking news headlines to predict the intraday direction of the USD-EUR pair in the foreign exchange market with an accuracy of 83.33%. Ferguson et al. [27] demonstrated that the long-short trading strategy with news sentiment has statistically significant daily risk-adjusted returns of 14.2 to 19 basis points. Chen et al. [28] applied genetic programming for performing dynamic proportion portfolio insurance and the approach showed promise over the traditional constant proportion portfolio insurance strategy. Mitra et al. [29] incorporated news sentiment in estimating equity portfolio volatility along with market information. Genetic programming has also been used in the area of technical trading, but has not been previously explored with analysis on sentiment. Healy and Lo [30] demonstrated a real-time news analytics framework to manage investment risks and returns with Thomson Reuters NewsScope data. Leinweber and Sisk [31] leveraged the predictability of market returns based on extracted news media sentiment and designed portfolio based trading strategies from sentiment signals.

The major contribution of this paper is to bridge the gap in the literature to develop a trading strategy with the use of sentiment feedback between news and tweets sentiment through genetic programming optimization. We argue that there is an opportunity to unravel the potential of their interaction effects because of the unique nature of the two information sources and their evident relationships with financial market movement. Using both sentiment sources, this study presents a novel framework for applying genetic programming method to optimize the performance of the trading strategy based on the sentiment indicator. The framework leverages existing empirical findings on the relationships observed among news sentiment, tweets sentiment and market returns. The key intuition behind the sentiment indicator is that the joint momentum of the two sentiment series leads to a robust signal for market anomalies which can be exploited in the form of above-average trading profits. For instance, if both news and tweets sentiments show strong momentum trending in one direction, the market return is likely to follow in the same direction. An investor can therefore establish a long position when the sentiment indicator generates such signal and exits when the reversal appears. In addition, the two information sources also display key distinguishable characteristics that the trading rules can be constructed by choosing the optimal trade-off between the speed of information release and the reliability of the information. In the study, we find that the sentiment indicator based genetic programming optimization approach yields a superior trading performance. The out-performance suggests that the sentiment-based indicator can be regarded as a valuable source of information and further validates the value of both news and tweets sentiment in exploiting trading opportunities. In addition, we conduct two experiments to evaluate the influence of trading costs on the profitability of the proposed trading strategies. The first experiment is to compute the break-even cost that eliminates the profits generated by the trading strategy. It captures the maximum cost percentage which the trading strategy can outperform the benchmark. The second experiment is a sensitivity test on trading costs. According to the empirical evidence, the trading costs of institutional investors on large-cap and liquid market is estimated to be 20 bps [32, 33, 34]. We set the trading costs to 10, 20 and 30 bps and test the profitability of proposed strategies. The results suggest that the sentiment feedback indicator generates robust profits and outperforms the benchmark under the consideration of trading costs.

The rest of the paper is organized as following. Section (2) introduces the sentiment feedback strength indicator and genetic programing methodology. Section (3) presents the three data sources of information along with their relationships, mainly in the form of tweets sentiment, news sentiment and market returns. Section (4) demonstrates the application of a dynamic and adaptive trading system with the proposed methodology. Section (5) discusses the key findings of the sentiment indicator based trading strategies under the genetic programming optimization framework and provides further explanation of the key findings. Section (6) concludes the discussion and points out some future research directions.

2. Methodology

Empirical evidence suggests that news sentiment, tweets sentiment and financial market returns are closely connected [9, 10, 13, 16, 17, 18, 24]. This section illustrates from an empirical perspective of the interactions among the three time series and how they can be exploited in the form of profitable trading opportunities. It is noted that the interaction is specific in unique time lag. For example, tweet sentiment and news sentiment are shown to elicit a lag-1 and lag-4 impact towards market returns respectively [18, 24] (see Figure 1). It is the hypothesis of this study that the empirical phenomenon of previous findings can be translated into a practical sentiment-based indicator that utilizes the concept of feedback strength, i.e. the joint momentum of the two sentiment series, to formulate profitable trading strategies.

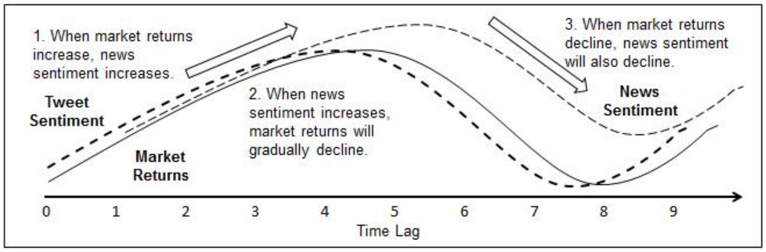


Figure 1: Relationships between tweets sentiment, news sentiment and market returns.

2.1. Tweets and news sentiment

For tweets sentiment, we adopt its definition from a former study by identifying a Twitter financial community and pinpoint its major influencers in the social network [18]. From a large-scale data crawling effort, we define a community as a group of relevant Twitter users with interests aligned with the financial market. We first identify 50 well-recognized investment experts' accounts in Twitter and use their common keywords to create the interests of the financial investment community. By constructing the two layers of the experts' followers, we apply a multitude of rigorous filtering criteria to establish a financial community boundary based on their persistent interests in the topic of financial investment [35]. After settling on a definition, we examine how messages from key influencers in the community interact with social mood or sentiment that tend to signal an impending upward or downward swing in the market price movement. We use key network metrics such as out-degree centrality (DC), betweenness centrality (BC) and closeness centrality (CC) to identify the financial community influencers and demonstrate that these key influencers along with their weight of influence in the financial community will provide better predictors of financial market movement measures [18]. We find that the BC group consistently outperformed the DC and CC groups. The sentiment regression model of the BC group has shown significance across all market returns at the level of 95%. Based on the empirical experiment, we adopt the use of the betweenness centrality in the model. In this study, we utilize the sentiment measure expressed by the key influencers in the Twitter financial community. The algorithm takes into consideration of the connectedness of the key influencers in the network, their sentiment scores and relevance of the message content benchmarked to a collection of financial entity words.

For extracting sentiment, we initially check words and phrases in each message with the financial entity word list, that the entities are scored based on their relevance to financial market [18]. The entity score of a message is defined as the highest score of all entities appeared in the message (see equation (1)).

$$\begin{aligned} &S_{entity}(i) = max(\omega(\mathbf{W^i_{message}} \cap \mathbf{W_{fe}})) \\ &i = 1, 2, \ldots, N \end{aligned} \tag{1}$$

where $S_{entity}(i)$ is the financial entity score of message i, $\mathbf{W^i_{message}}$ is the word set split from message i, $\mathbf{W_{fe}}$ is the financial entity word set, $\omega(\mathbf{W})$ is the financial entity weight set of word set $\mathbf{W}$, N is the number of message and $n_{matched}$ is the number of matched financial entities.

Our sentiment algorithm is based on the use of the SentiWordNet[2] dictionary, a lexical resource with words linked to sentimental scores. SentiWordNet assigns corresponding sentiment scores to word entities in terms of positivity, negativity, and objectivity [36]. The SentiWordNet dictionary has been widely applied in recent studies with sentiment analysis [37, 38, 39, 40]. In this study, we compute the message sentiment score as the average of SentiWordNet sentiment for all words in the message, with the adjustment of financial entity score (see equation (2)).

$$\begin{aligned} &S_{sentiment}(i) = \frac{\sum_j n_i^j \times s(j)}{\sum_j n_i^j} \times S^i_{entity} \times \mathrm{sgn}(i) \\ &\mathrm{sgn}(i) = \begin{cases} -1 & \text{if } \mathbf{W^i_{message}} \cap \mathbf{W_{neg}} \neq \emptyset \\ 1 & \text{others} \end{cases} \end{aligned} \tag{2}$$

[2]The SentiWordNet dictionary a lexical resource explicitly devised for supporting sentiment classification and opinion mining applications, and it is available on http://sentiwordnet.isti.cnr.it/.

where $S_{sentiment}(i)$ is the sentiment score of message i, $S_{entity}(i)$ is the financial entity score of message i computed by Eq.(1), $\mathbf{W^i_{message}}$ is the word set split from message i, $\mathbf{W_{neg}}$ is the negative connotation word set, n_i^j is the number of occurrence of SentiWordNet word j in message i, $s(j)$ if the sentiment score of word j.

According to our previous findings that Twitter user centrality determines the influence of their messages in the financial community, we include the user centrality score in the daily tweets sentiment calculation. The daily user sentiment measure is the average score of all messages in each day (see equation (3)), which leads to the computation of the daily tweets sentiment as the weighted average user sentiment score (see equation (4)).

$$S_{user}(i,t) = \frac{\sum_{k=1}^{n_i^{(t)}} S_{sentiment}(k)}{\sum_{k=1}^{n_i^{(t)}} S_{entity}(k)} \tag{3}$$

where $S_{user}(i,t)$ is the daily user sentiment score of user i on day t, $n_j^{(t)}$ is the number of message by user j on day t, $S_{sentiment}(k)$ and $S_{entity}(k)$ are sentiment score and entity score of message k.

$$S_{tweets}(t) = \frac{1}{N}\sum_{j=1}^{N} \omega_j S_{user}(i,t) \tag{4}$$

where $S_{tweets}(t)$ is the daily tweets sentiment score on day t, ω_j is the centrality score of user j.

For news sentiment, we also follow the lexicon-based approach in leveraging the word dictionaries to generate sentiment scores. Through a four-step procedure, we convert the raw text format into daily news sentiment score for the empirical study. With the complex textual structure, we initially decompose the raw text into individual words with the removal of stop words. We then apply lemmatization techniques to convert different inflicted forms of a word into a uniform entity. For instance, we would regard “rising”, “risen” and “rises” as the word entity “rise”. For each word in the news content, we extract the associated score from the sentiment dictionary and finally, we generate the sentiment score for each news text by averaging all individual word scores. To compute the daily news sentiment, we aggregate all news articles published in each day and compute the daily average value of news sentiment scores (see Eq. 5). In addition, the relative publication frequency of individual vendors is accounted for in the calculation. The intuition is that the more established media entities tend to publish more frequently and therefore, their news articles can reach a wide audience.

$$S_{news}(t) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_i}\sum_{j=1}^{n_i}S(j) \tag{5}$$

where $S_{news}(t)$ is the daily news sentiment score on day t, $S(j)$ is the SentiWordNet sentiment score for word j, n_i number of word in new article i, N is the total number of news article in a day.

2.2. Sentiment feedback strength indicator

We construct a sentiment indicator based on the feedback strength of two sentiment series. The intuition behind the feedback strength indicator is that the joint momentum of the two sentiment series leads to significant market anomalies which can be exploited in the form of above-average trading profits. The feedback relation between the sentiment series can be a valuable source of information to explain market movement. For instance, if both news and tweets sentiments show strong momentum in trending in one direction, the market return is likely to follow in the same direction. An investor can therefore establish a long position when the sentiment indicator generates such signal and exits when the reversal appears. Aligned with the intuition to reflect the joint momentum of the sentiment series, a weighted scoring approach is applied with the weight of respective sentiment.

$$\begin{cases} SMA^{n}_{news}(t) = \frac{1}{n}\sum_{i=t-n}^{t-1} S_{news}(i) \\ SMA^{n}_{tweets}(t) = \frac{1}{n}\sum_{i=t-n}^{t-1} S_{tweets}(i) \end{cases} \tag{6}$$

where $SMA^{n}_{news}(t)$ and $SMA^{n}_{tweets}(t)$ are the simple moving average of the news sentiment series and tweets sentiment series with a window of n periods at time t, $S_{news}(i)$ and $S_{tweets}(i)$ are the value of the news sentiment and tweets sentiment at time i respectively.

$$FSI^{(\omega_{news},\omega_{tweets},n_{news},n_{tweets})}(t) = \omega_{news} SMA_{news}^{n_{news}}(t) + \omega_{tweets} SMA_{tweets}^{n_{tweets}}(t) \tag{7}$$

where the $FSI(t)$ is the feedback strength indicator with the linear function of simple moving averages for news and tweets sentiment, ω_{news} and $\omega_{tweets} = 1 - \omega_{news}$ are the weights of news sentiment and tweets sentiment, n_{news} and n_{tweets} are moving average periods for news and tweets respectively.

2.3. Technical indicators

In addition to the sentiment feedback strength indicator, the moving average convergence/divergence (MACD) and the relative strength index (RSI) are chosen as the technical indicators. The two indicators have been frequently referenced in existing technical trading literature with substantial value in forecasting market direction [41, 42]. Fang et al. [43] suggested the predictive power of these two indicators with an evolutionary trend reversion model. The inclusion of these technical indicators expands the search space in the genetic programming framework with our sentiment feedback strength indicator. The formulas of the respective technical indicator are defined as below:

- Moving Average Convergence Divergence (MACD)

  The Moving Average Convergence Divergence (MACD) was first invented by Gerald Appel and later enhanced by Thomas Aspray [44, 45]. The intuition behind the MACD indicator is that the comparison between the short- and long-term moving averages of an underlying stock's movement plays a significant role in signaling short-term price momentum. The indicator has been widely applied to identify the trend direction and momentum.

$$EMA_p^n(t) = \frac{2}{n+1} p(t) + \frac{n-1}{n+1} EMA_p^n(t-1) \tag{8}$$

where $EMA_p^n(t)$ is the exponential moving average of market price with a window of n periods at time t, $p(t)$ is the market price at time t.

$$MACD_p^{(n_{long},n_{short})}(t) = EMA_p^{n_{long}}(t) - EMA_p^{n_{short}}(t) \tag{9}$$

where $MACD_p(n_{long}, n_{short})$ is the MACD between the short-term and long-term moving average of the market price with n_{short} periods and n_{long} periods respectively.

- Relative Strength Index (RSI)

  The Relative Strength Index (RSI) was developed by J. Welles [46]. The intuition behind the RSI indicator is its evaluation of the current and historical strength of the stock within a recent trading period. RSI is regarded as a momentum oscillator that measures the magnitude and speed of the underlying stock's price movement. It has been identified to identify trends, divergence and overbought and oversold conditions [46].

$$RS^n(t) = \frac{SMA_{p_{up}}^n(t)}{SMA_{p_{down}}^n(t)} \tag{10}$$

  where $RS^n(t)$ is the relative strength on day t during with n days window, $SMA_{p_{up}}^n(t)$ and $SMA_{p_{down}}^n(t)$ are the average up prices and the average down prices during the past n days.

$$RSI^n(t) = 100 - \frac{100}{1 + RS^n(t)} \tag{11}$$

  where $RSI^n(t)$ is the relative strength indicator of day t.

2.4. Genetic programming as an optimization approach

Genetic programming is a special class of genetic algorithm, which was first developed by John Holland in 1992. Genetic algorithm was built on the premise of the natural selection process that individual action with condition is evaluated with a pre-specified fitness function until the optimal combination is reached. Holland illustrated that "a population of fixed length character strings can be genetically bred using the Darwinian operation of fitness proportionate reproduction and the genetic operation of recombination" [47]. The central goal of using genetic algorithm is to exploit a vast region in the search space and at the same time to manipulate variations of strings [48]. The difference between genetic programming and genetic algorithm lies on the representation of the varying string length in the search space. Genetic programming is an iterative algorithm that searches for optimal program with the objective of satisfying the best fitness function (see Algorithm 1). It allows solutions to be represented by a flexible string length with the Boolean operators connecting the combinations of indicators (see Figure 2). For example, we can construct solutions with different combinations of indicators and parameters in contrast to the fixed set of indicators that we have to use for each search. Moreover, GP requires input solutions to be represented in a tree structure to accommodate the flexibility. Three major genetic operators are applied to a given problem during the optimization process: mutation, crossover and encoding.

1: Randomly create an initial population of individuals from the available function and terminal set;

repeat

> **2:** Execute each individual and compute its fitness;
>
> **3:** Select one or two individual(s) from the population with a fitness-based probability to participate in genetic operations (i.e. crossover and mutation). ;
>
> **4:** Create new individual(s) by applying genetic operations with specified probabilities of crossover or mutation;

until *Stopping condition is met*;

return the best-so-far individual;

Algorithm 1: Genetic Programming algorithm

2.4.1. Function and terminal set

One of the preparatory steps in genetic programming is to specify a set of functions and terminals, that are essential to establish a diverse universe of programs in the search space. A function set represents as the branches within the tree structure and typically includes statements, operators and functions [49]. In our study, the boolean operators, sentiment feedback strength and technical indicators constitute the primary function set in the optimization framework. A terminal set, on the other hand, represents the parent nodes in the tree structure and is designed to parameterize the specified function set. As one of its key forms suggested by Banzhaf et al. [49], constants are selected as terminal set in the form of integer, real number and boolean variable.

Function Set:

– Trading Signal Functions

1. Sentiment Strength Indicator Signal

$$Signal_{FSI}^{(\omega_{news},\omega_{tweets},n_{news},n_{tweets},\theta)}(t) = \begin{cases} 1, & FSI^{(\omega_{news},\omega_{tweets},n_{news},n_{tweets})}(t-1) > \theta \\ 0, & otherwise \end{cases} \tag{12}$$

where $Signal_{FSI}^{t}(\omega_{news},\omega_{tweets},n_{news},n_{tweets},\theta)$ is the binary feedback strength signal with threshold of the summation θ.

2. MACD Signal

$$Signal_{MACD}^{(n_{long},n_{short},n_{signal})}(t) = \begin{cases} 1, & MACD_{p}^{(n_{long},n_{short})}(t-2) < EMA_{MACD}^{n_{signal}}(t-2) \\ & \& MACD_{p}^{(n_{long},n_{short})}(t-1) > EMA_{MACD}^{n_{signal}}(t-1) \\ 0, & MACD_{price}^{(n_{long},n_{short})}(t-2) > EMA_{MACD}^{n_{signal}}(t-2) \\ & \& MACD_{price}^{(n_{long},n_{short})}(t-1) < EMA_{MACD}^{n_{signal}}(t-1) \\ & Signal_{MACD}^{(n_{long},n_{short},n_{signal})}(t-1), Others \end{cases} \tag{13}$$

where $Signal^{t}_{MACD}(n_{long}, n_{short}, n_{signal})$ is the binary signal of MACD crossover.

3. RSI Signal

$$Signal^{(n,\theta)}_{RSI}(t) = \begin{cases} 1, & RSI^{n}(t-2) < \theta \\ & \&RSI^{n}(t-1) > \theta \\ 0, & RSI^{n}(t-2) < (100-\theta) \\ & \&RSI^{n}(t-1) > (100-\theta) \\ Signal^{(n,\theta)}_{RSI}(t-1), & Others \end{cases} \tag{14}$$

where $Signal^{(n,\theta)}_{RSI}$ is the binary signal of RSI indicator.

– Boolean Operator: AND, OR

Terminal Set:

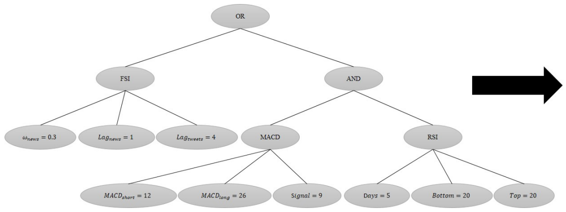
– Integer constants: discrete uniform random variable from $\{n \in \mathbb{N} : 1 \leq n \leq 30\}$

– Real constants: uniform random variable from $\{r \in \mathbb{R} : -0.1 \leq n \leq 0.1\}$

– Boolean constants: True, False

2.4.2. Representation

This study applies the standard framework of genetic programming to locate the optimal trading strategy with the proposed sentiment indicator based and technical indicator based trading rules. Based on our design, each individual represents a trading strategy with combination of different trading rules and optimized parameters. In the genetic programming (GP) framework, we first initialize the population of programs constructed from the sentiment feedback strength indicator and the technical indicators. Through the search process, we incorporate the Boolean operators, "AND" and "OR", for allowing different combinations of the indicators. For instance, the framework has the ability to develop a trading strategy with the configuration in the form of:



Date	FSI	MACD	RSI	Signal
01/02/2013	TRUE	FALSE	TRUE	TRUE
01/03/2013	TRUE	FALSE	FALSE	TRUE
01/04/2013	FALSE	TRUE	FALSE	FALSE
...	...	...	...	...

Figure 2: Sample tree structure outputs of Genetic Programming.

$FSI(0.3, 1, 4, 0.25) \| (MACD(12, 26, 9) \& RSI(14, 20))$ (see Figure 2). As demonstrated, the sample strategy is a resultant combination of three indicators with its corresponding parameters. As a natural extension, the algorithm then generates trading signals in the form of "TRUE/FALSE" signal at each time period and computes its respective trading performance (see Figure 2). Following the design of long-only strategy, "TRUE" and "FALSE" signal represent long position and empty positions respectively. For a "TRUE" signal, the system records the cumulative returns over the holding period until a reversal of the trading signal appears. For example, if a position is established on day 1 and closed on day 10, the trading return is calculated as the cumulative returns over the 10-day period. With the trading signals generated by the strategy, the algorithm ascertains its fitness and then performs genetic operations in crossover and mutation with specified probabilities.

2.4.3. Fitness evaluation: Sterling ratio

Risk-adjusted measures are commonly used when evaluating trading strategies. Under the classical framework of the Capital Asset Pricing Model (CAPM), Sharpe ratio, Jensen's alpha, Treynor's ratio and Information ratio were first proposed. The focus to prevent substantial loss has then become a priority for investors, and risk-adjusted measures, such as Sterling ratio, Calmar ratio and Burke ratio, have been designed to factor drawdown into the underlying risk measures. Over time, there are more advanced measures for assessing the performance of trading strategies such as the lower partial moment, Sortino ratio, Kappa measures, Value at risk (VaR) along with its alternative conditional and modified forms (CVaR and MVaR). Over considerations of the feasibility and compatibility for the sentiment indicator based genetic programming framework, we select Sterling ratio as the performance measure for evaluating the profitability of the trading strategy population. The reasons are twofold: First, Sterling ratio captures the total cumulative returns over the duration in the trading strategy which is often the most important criteria for measuring financial success. Second, the ratio takes into consideration of the maximum drawdown over a specified investment period which investors are often more sensitive over other risk measures. The advantage of using the maximum drawdown over standard deviation is the emphasis on downside risk that the trading strategy does not yield substantial losses and therefore achieve capital protection.

$$Sterling = \frac{R_{total}}{1 - avg(max(Drawdown))} \tag{15}$$

where $Sterling$ is the Sterling ratio for one strategy, R_{total} is the total cumulative returns and $avg(max(Drawdown))$ is the average monthly maximum drawdown.

2.4.4. Genetic operators

Mutation and crossover are used as the primary genetic operators under the genetic programming framework. These genetic operations facilitate the evolutionary process by generating individual programs for the new population. As a result, they increase the speed of convergence and the likelihood of the optimal solution achieving global optimal point. Both parameters in the genetic operators are determined based on a sensitivity assessment of the genetic programming framework. We observe that the computational efficiency is severely impacted when the selection criterion is too aggressive. On the other hand, conservative parameter selection leads to suboptimal convergence, possibly due to omission of significant changes within the tree structure.

- Crossover is the process when two parent individuals combine to generate new offspring. In other words, the parents swap subtrees from each other

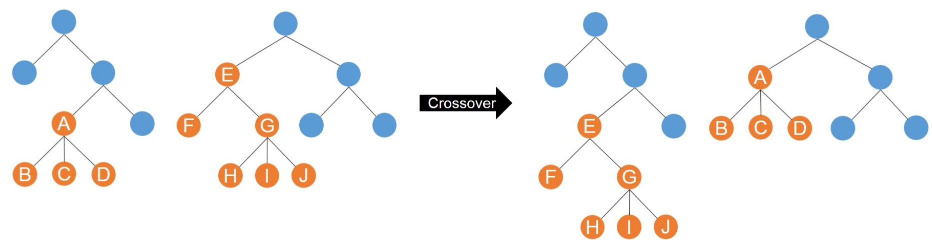


Figure 3: Crossover diagram.

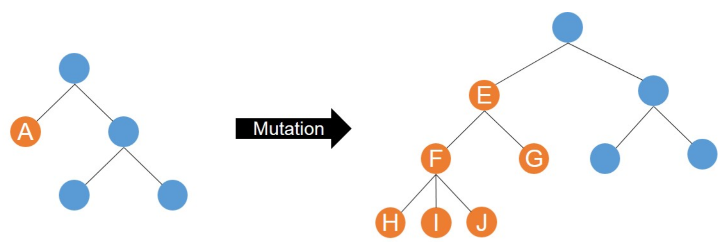


Figure 4: Mutation diagram.

and form new individuals (see Figure 3). The purpose of the crossover operation is to explore combinations of good-performing individuals to form larger and better individuals. In our design, each node (along with its subtree) in the parent individuals is set to follow a 50% probability of crossover.

- Two types of mutation can be applied in genetic programming: function replacement and subtree replacement. First, the first type retains the original tree structure by replacing one node while the second type replaces the whole subtree stemmed from one node. Due to the complexity of functions and type constraints in our framework, we prefer the subtree replacement as the mutation mechanism with a 90% probability for each node in an individual program (see Figure 4).

3. Data

Using Twitter API, we collect a total of $1,271,308$ tweet messages between August 1, 2012 and January 30, 2015 from a selective group of Twitter users. These users, known as critical nodes, are situated at the most central locations among the Twitter Financial Community which we define as a subset of the Twitter universe with direct interest and relevance to the financial market [35, 18]. The empirical evidence suggests that the sentiment expressed by these critical nodes has predictive and persistent relationship with key financial market indices [18]. For this study, we select the top 200 users with the highest betweenness centrality which was determined to provide the most significant signal on explaining market returns. Our previous study found that selecting the top 200 users in the betweenness centrality group provides the most significant signal, and a group with more than 200 critical node users dilutes the significance of the model but the result remains robust at a high significance level[18]. To address over-fitting problem, these influences will be reevaluated during the training period and the influential accounts are independently reselected based on the set criteria.

We build a news crawler that extracts relevant market-related news entries from the Northern Light SinglePoint business news portal and pre-processes them into news sentiment. The news crawler features a Java-based platform that utilizes pre-specified query such as S&P 500 and NASDAQ, and records attributes such as the title, summary, description and the sentiment. The news data contains 569 business days across the evaluation period. A total of 2,420 distinct news providers were captured within this dataset, representing a diverse group of media sources. Textual information collection based on query searching is a widely used approach. Differing from most research that used a list of market-related queries to gather information that represents the overall market condition [50, 51], we use query to collect news related to specific stock tickers. For example, the stock ticker “C” is linked with query “Citi Group” so that news content containing this query will be stored and matched with the ticker.

In summary, a total of $678,378$ news articles that are collected between August 1, 2012 and January 30, 2015 are extracted from the database for the empirical analysis.

The S&P 500 ETF is used in this study for analysis, whose characteristics largely mirror the underlying S&P 500 index in terms of price and yield performance. Given its wide popularity among institutional and retail investors, we choose the S&P 500 ETF as a suitable representation of the U.S. broad market performance in this study. We collect daily historical return of these indices through Bloomberg Terminal from July 31, 2012 to January 30, 2015. The data was based on end-of-day price (closing price) for the U.S. domestic market. An important pre-processing step is to transform the index prices into log-returns and align them with corresponding tweets and news sentiment.

Market movement has been empirically linked to tweet and news sentiment in previous studies. First, we show that the tweet sentiment expressed by the critical nodes has a significant lag-1 relation with major market indices including the S&P 500 index [18] (see Table 7). It illustrates that the community would be more representative to market participant's beliefs, and consequently the sentiment extracted from this financial community would serve as a better predictor to the market movement. Second, news sentiment exhibits a lag-4 effect on market returns and conversely market returns elicit consistent lag-1 and lag-2 effects on news sentiment [24] (see Table 8). This finding suggests that news sentiment drives trading activity and investment decisions. Subsequently, heightened investment activity further stimulates involuntary responses, which manifest in the form of more news coverage and publications.

In addition, we would like to investigate whether news and tweet sentiments have relation with each other in different time scales. As a novel empirical finding, the results of the linear regression model reveals a significant lag-3 impact from news sentiment to tweet sentiment (see Table 9). In other words, the news sentiment from three days ago helps explain the tweet sentiment on the current day. This finding reaffirms the previous studies on the lag-1 and lag-4 impact from tweets and news sentiment respectively. Furthermore, it

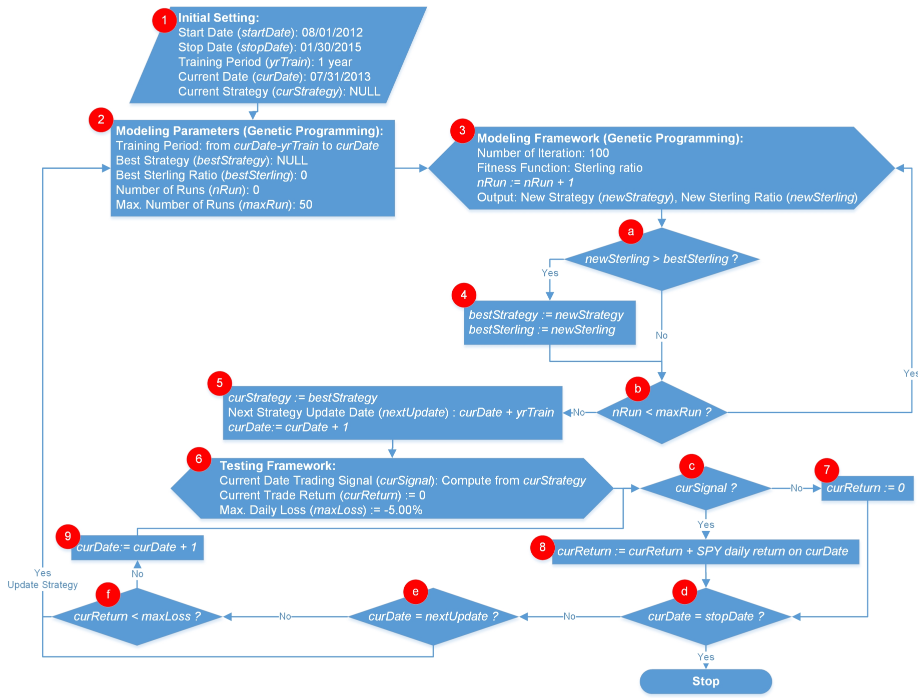


Figure 5: Trading system flow chart.

unravels the potential interplay between the two sentiment series and provides an intuitive basis for formulating and validating the sentiment indicator based trading strategy.

4. Application: Trading system

4.1. System overview and construction

The objective of this section is to showcase the existence of profitable trading opportunities using both price and sentiment dataset. With the proposed genetic programming methodology, we extend its application and construct a dynamic and adaptive trading system. The motivation of such system hinges on the realistic need of a practical trader who not only expects positive risk-adjusted returns, but also demands its configuration to evolve with changing market conditions. Therefore, the design of the trading system emphasizes on three main features: performance, risk mitigation and dynamic adaptation. The first two features aim at generating superior performance with absolute financial gains but avoiding substantial loss scenario. The last feature is the dynamic capability of the trading system to adjust its model parameters with more recent data set.

To further illustrate the construction process, the trading system can be decomposed into two frameworks in modeling and testing (see Figure 5). The entire process is adaptive to both the training period and the maximum loss. If the training data is more than one year old, the system will automatically retrain the model with the updated data. If the trading loss exceeds a preset stop-loss threshold, the system will also automatically retrain the model with the most updated data. The parallelogram represents the initial setup parameters; the rectangle boxes represent major computational steps; the rhombus boxes represent major decision points; the two major components are denoted as step 3 and 6.

4.1.1. The modeling framework

The design of the modeling framework utilizes the genetic programming methodology in constructing the underlying model with the training data. In this study, the initial setting is first defined and the modeling parameters are used to govern the specification of the optimization framework (see step 1 and step 2 in Figure 5). The duration of the training period is selected as one year, which the system takes the first year of data for model training and reserves the subsequent year of data in the testing framework. With the specified indicator set, the aim of this framework is to identify the strategy with the best performance in Sterling ratio through an iterative process. As for the details of the experimentation, we allow 100 iterations for each run of the experiment to test whether a strategy yields a better Sterling ratio compared to its predecessor (see step 3 in Figure 5). The modeling framework continues to search for the best strategy until the maximum number of runs is reached (i.e. 50 runs), and both configuration of the strategy and its corresponding Sterling ratio are recorded (see step a and step b in Figure 5).

4.1.2. The testing framework

With the best strategy identified from the modeling framework, the trading system tests its performance against two conditions for potential strategy adjustment (see step 4 to step 9 in Figure 5). First, the system actively checks whether the underlying strategy results in a daily loss of more than a threshold percentage during the testing period. If the condition is met, the system performs an adjustment to the strategy by running the modeling framework with the training data dating back from the time when the major loss is incurred (see step f in Figure 5). Second, there is a time condition applied to limit the duration of the out-of-sample period as time evolves (see step e in Figure 5). In the testing framework, the duration of the out-of-sample period is set to one year. In other words, the system utilizes the subsequent year of data following the training period. Both conditions allow the trading system to continuously adapt to changing market landscape with a more robust and risk-averse underlying strategy.

4.2. Experimentation for trading system

One of the important milestones of the study is to evaluate the performance of the trading strategies with different groups of indicators. Using the trading system, we establish three independent function sets with unique groups of indicators in the genetic programming configuration:

1. Combination of Sentiment indicator and Technical indicators
2. Sentiment indicator only
3. Technical indicator only

This arrangement seeks to distinguish the difference among indicator groups and therefore yields insight towards the value of using a sentiment based indicator. Table 1 lists the trading system parameters with respect to the three corresponding systems.

Table 1: Trading system parameters.

		Combination	Sentiment-only	Technical-only
Initialization	Data	08/01/2012-01/30/2015		
	Training Period	1 Year		
	Testing Period	1 Year		
Genetic Programming Parameters	Function Set	FSI, MACD, RSI	FSI	MACD, RSI
	Crossover Rate	0.5		
	Mutation Rate	0.9		
Testing Parameter	Maximum-Loss	-5.00%		

4.2.1. *Trading strategy benchmark: Buy-and-hold strategy*

For additional comparison, we establish a buy-and-hold strategy as the benchmark for evaluating the performance of the trading strategy. At the beginning of the holding period, an open position is established by buying the S&P 500 ETF and the unrealized profit and loss is recorded on a daily basis. At the end of the period, the position on S&P 500 ETF is closed. In the study, we use Sterling ratio, Total Profit/Loss, Average Profit/Loss per trade, Standard Deviation, Percentage of Winning Trades and Trading % as the major performance measures in the evaluation process.

4.2.2. *Trading costs*

Trading costs can be broadly categorized in terms of explicit costs such as brokerage and taxes, and implicit costs, which include market impact costs, price movement cost, and opportunity cost, etc. The implicit costs normally have to be estimated. Considering that trading costs affect the profitability of the trading strategies, we implement two additional tests to evaluate the chance of proposed trading system to survive from both implicit and explicit impacts.

First, we estimate the break-even costs following the "double-or-out" strategy according to Bessembinder and Chan [52]. In this experiment, the trader borrows capital to hold two positions when the strategy is "in market" and holds one standard long position when the strategy is "out of market" [52, 53, 54].

A larger break-even cost indicates a stronger ability to tolerate market impacts and a higher flexibility regarding the trading timeliness. In the second experiment, we apply sensitivity analysis by testing a series of trading costs in the system. This is useful to evaluate the impacts to dynamic readjustments and trading decisions. Several empirical studies have evaluated trading costs in the U.S. market. Chan and Lakonishok [32] designed the trading cost evaluation based on market capitalization and trade complexity. They illustrated that the average round-trip total costs for large-cap stock trades on NYSE is 20 bps. In a similar study, Keim and Madhavan [33] decomposed trading costs into explicit and implicit components and argued that the average total costs in NYSE and AMEX range from 30 to 200 basis points. For pension fund, the market impact cost and execution costs for buy orders are estimated as 20 basis points and 27 basis points respectively [34]. The choices of trading costs are functions of the underlying trading securities and markets. For example, Fong and Yong [53] chose 50 bps according to the analysis on global equity market, and Harris and Yilmaz [55] chose a lower rate of 10 bps for foreign exchange market. In this study, we test the trading system performance with one-way trading costs of 10 bps, 20 bps, and 30 bps that are consistent with the empirical evidences.

5. Key findings

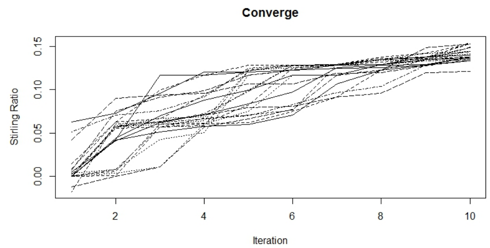
The key findings of the study can be decomposed into two parts. The first part presents the statistical evidence of out-performance using the genetic programming methodology with the proposed sentiment feedback strength indicator. To further confirm its validity, the second part demonstrates the existence of a profitable trading strategy through the construction of a dynamic and adaptive trading system, through its superior comparison of performance against other established benchmarks.

5.1. Optimization findings

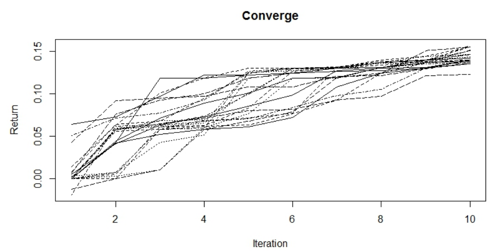
To evaluate the performance of the genetic programming algorithm, we conduct 1,000 experiments with each indicator group as the primary function set:

sentiment-indicator-only, technical-indicator-only, and the combination of both indicators. For each experiment, the number of iterations is chosen as 100 times to ensure better convergence and consistency of the solution. In addition, the algorithm performance is compared using both full period (08/01/2012-01/30/2015) and out-of-sample period (11/01/2013-01/30/2015). The optimization results suggest that the sentiment-indicator-only strategy is superior to the combination approach and the technical-indicator-only strategy in terms of Sterling ratio and the total return. The finding is also significant for both full-period and out-of-sample period. Another noted observation is that the winning percentage is among the highest for the sentiment-indicator-only strategy at an average of 58%. In terms of the dispersion of the results, the sentient-indicator-only strategy and the combination approach exhibit the highest consistency with smaller inter-quantile range . The technical-indicator-only strategy, on the other hand, yields a wider range of results suggesting that its performance is not as reliable as the other two experiments. In addition, it is crucial for the genetic programming framework to converge towards the optimal solution within the search space. We record the corresponding returns, risk, Sterling ratio and the number of trades for each iteration. The findings show that all 4 measures converge within 10 iterations, suggesting that the genetic programming framework provides reliable and efficient output (see Figure 6).

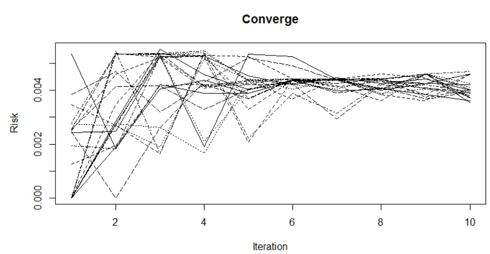
To further validate the robustness of the above results, we conduct the Analysis of Variance to examine whether there is any statistical difference among the means of the three optimization results. Comparing Sterling Ratio, Total Profit, Winning Percentage and Sharpe Ratio, we find that the performance metrics differ significantly among different indicator groups (See Table 2). Moreover, this finding is consistent across the out-of-sample period with p-value under 0.05. As an extension to reveal information about the relative difference, the Tukey's Honest Significance test is conducted at the confidence level of 95%. The result suggests that the sentiment-indicator-only strategy outperforms the combination approach and the technical-indicator-only strategy in terms of all four performance metrics with a significance level under 0.05 (see Table 3). The



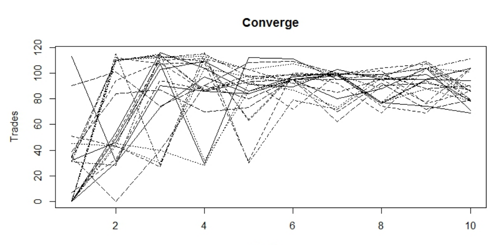
(a) Sterling ratio convergence



(b) Returns convergence



(c) Risk convergence



(d) Number of trades convergence

Figure 6: Convergence of Sterling ratio, returns, risk, and number of trades.

technical-indicator-only strategy yields the worst performance, followed by the combination approach. The result using the Tukey's test is also consistent using the out-of-sample period (see Table 4).

Table 2: ANOVA test of different performance measures

	Full Period		Out-Sample Period	
	F-statistic	p-value	F-statistic	p-value
Sterling Ratio	168.2	$< 2 \times 10^{-16***}$	562.4	$< 2 \times 10^{-16***}$
Total Profit	225.7	$< 2 \times 10^{-16***}$	591.1	$< 2 \times 10^{-16***}$
Winning %	1111	$< 2 \times 10^{-16***}$	734.8	$< 2 \times 10^{-16***}$
Sharpe Ratio	20.86	$1.01 \times 10^{-9***}$	441.6	$< 2 \times 10^{-16***}$

Note: *** 1% confidence level,** 5% confidence level, and * 10% confidence level.

Table 3: Tukey's Honest significance test (Full Period)

	"SI+TI" vs. "SI"		"TI" vs. "SI"		"TI" vs. "SI+TI"	
	Diff.	p-value	Diff.	p-value	Diff.	p-value
Sterling Ratio	-0.56%	0.00***	-1.20%	0.00***	-0.65%	0.00***
Total Profit	-2.35%	0.00***	-5.13%	0.00***	-2.79%	0.00***
Winning %	-1.48%	0.00***	-3.91%	0.00***	-2.43%	0.00***
Sharpe Ratio	-0.038	$1.00e-07^{***}$	-0.039	$1.00e-07^{***}$	0.001	0.996

Note: *** 1% confidence level,** 5% confidence level, and * 10% confidence level.

Table 4: Tukey's Honest significance test (Out-Sample Period)

	"SI+TI" vs."SI"		"TI" vs. "SI"		"TI" vs. "SI+TI"	
	Difference	p-value	Difference	p-value	Difference	p-value
Sterling Ratio	-1.16%	0.00***	-3.33%	0.00***	-2.17%	0.00***
Total Profit	-1.67%	0.00***	-4.83%	0.00***	-3.16%	0.00***
Winning %	-1.08%	0.00***	-3.06%	0.00***	-1.99%	0.00***
Sharpe Ratio	-0.092	0.00***	-0.262	0.00***	-0.170	0.00***

Note: *** 1% confidence level,** 5% confidence level, and * 10% confidence level.

5.2. Trading system performance comparison

Using the trading system, this section presents the performance comparison of the sentiment feedback strength based trading strategies against two benchmarks. The first benchmark is a strategy that utilizes the genetic programming framework to generate trading signals based on entirely technical indicators only. The rationale behind this strategy is that GP can generate useful technical trading rules with optimal set of parameters. The second benchmark is the traditional buy-and-hold strategy that is commonly utilized by small investors and mutual funds.

The comparison results show that both sentiment indicator based trading strategies provide a clear edge over the two benchmark strategies in terms of higher Sterling ratio and total profit/loss (see Figure 7). The optimal sentiment-only strategy and combination approach generate over 14.7% Sterling ratio compared to 10.4% and 13.6% from technical indicators-only strategy and the buy-and-hold strategy respectively (see Table 5). For the comparison of the total profit/loss over the evaluation period, the sentiment feedback strength based strategies yield the best performances at cumulative returns over 25.5% compared to 17.4% and 23.3% from the two benchmark strategies respectively. On the other hand, the results related to the sentiment indicator based trading strategies suggest that sentiment provides support in controlling loss indicated by the significantly lower monthly maximum drawdown at -5.9% and -7.2% in contrast to -8.1% and -7.6% for the two benchmark strategies. We find that the percentage of winning trades is also higher at 57.0% and 57.6%. From a standpoint of evaluating the strategy risk, the standard deviation of the daily returns is slightly lower at 10.7% and 10.8% compared to 11.2% and 11.3% for the technical indicator strategy and the buy-and-hold strategy respectively.

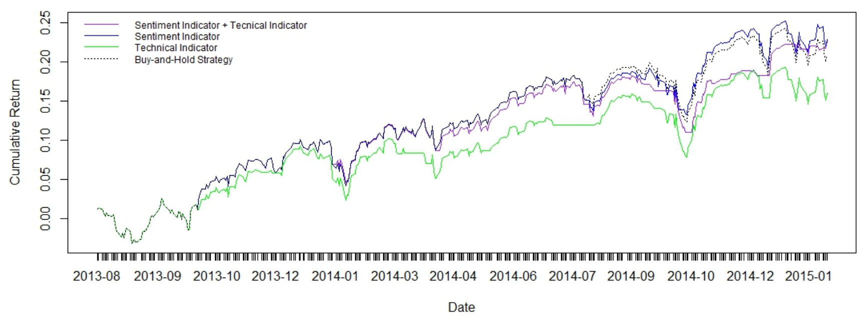


Figure 7: Cumulative return of trading strategies.

Table 5: GP optimization trading strategy performance

	Sentiment Indicator	Combination	Technical Indicators	Buy-and-Hold Strategy
Number of Testing Days	377	377	377	377
Percentage of Winning	57.0%	57.6%	54.4%	56.8%
Total Profit/Loss	25.6%	25.5%	17.4%	23.3%
Standard Deviation	10.7%	10.8%	11.2%	11.3%
Monthly Max Drawdown	-5.9%	-7.2%	-8.1%	-7.6%
Sterling Ratio	14.8%	14.7%	10.4%	13.6%
Sharpe Ratio	1.36	1.41	0.96	1.24

5.3. Trading profits and trading costs

This section examines the impact of trading costs on the profitability of trading strategies. During the estimation process, it is our goal to come up with a trading cost that is as realistic and reasonable as possible. In the evaluation of break-even cost, we do not consider the technical-only strategy as its profit is lower than the S&P500 benchmark. For sentiment-only strategy and combination strategy, the round-trip break-even costs are found to be 92 bps and 10 bps respectively. Although the two strategies exhibit similar profitability without the consideration of trading costs, the higher break-even cost for sentiment-only strategy indicates that the strategy can endure larger market impacts and still be more profitable than combination strategy in real transactions. In addition, empirical evidences showed that the average round-trip trading cost of large-cap stocks on NYSE is at least 20 bps [32, 33], which is significantly lower than that for the sentiment-only strategy. We do not consider margin cost as the study focuses on a long-only strategy, which does not require borrowing money to purchase stock. In addition, we do not evaluate the impact of order size on the performance comparison due to the observation that S&P500 is a highly liquid market. With transaction costs and market impacts, the sentiment-only strategy yields superior risk-adjusted return ratio while the combination-strategy does not.

In the sensitivity test, we applied one-way trading costs of 10 bps, 20 bps, and 30 bps respectively. In these experiments, the profitability of sentiment-only strategy is largely unaffected. In particular, when the trading cost is less than 20 bps, the sentiment-only strategy outperforms the buy-and-hold strategy using the S&P 500 market benchmark (see Table 6). A noted observation is that, in this test, the threshold of eliminating all profits over benchmark turns out to be around 20 bps, significantly lower than the break-even cost. The rationale is that the extra costs trigger the loss readjustment earlier and then generate different trading flows. Overall, the profitability of sentiment-only strategy is acceptable for institutional investors based on the assumption of less than 30 bps round-trip trading cost. The combination strategy fails to keep the high level of profits under the lowest 10 bps setting as it generates excessive turnover. For example, the number of trades for sentiment-only strategy and combination strategy with 20 bps trading cost are 7 and 14 respectively. Through this sensitivity test, we demonstrate that the sentiment-indicator strategy provides the best trade-off between capturing price trend and over-trading.

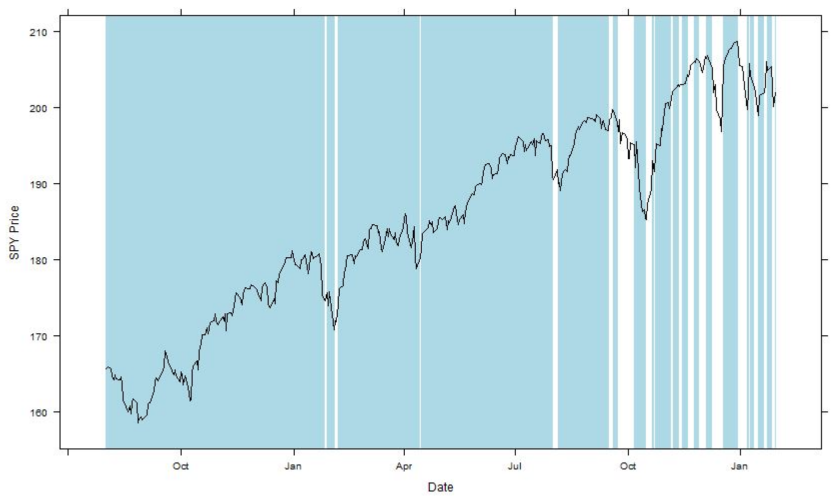


Figure 8: Optimal sentiment indicator based trading strategy signals.

Note: The line shows the S&P 500 ETF price, and the background shaded areas indicate the holding periods.

Table 6: Trading strategy performance with trading costs

Trading Cost (One-Way)	Performance Measure	Sentiment Indicator	Combination	Technical Indicators
10 bps	Total Profit/Loss	24.7%	20.2%	12.1%
	Sterling Ratio	14.6%	12.1%	7.5%
	Sharpe Ratio	1.33	1.21	0.81
20 bps	Total Profit/Loss	23.0%	17.4%	8.9%
	Sterling Ratio	13.6%	10.6%	5.6%
	Sharpe Ratio	1.24	1.06	0.61
30 bps	Total Profit/Loss	21.3%	14.8%	5.8%
	Sterling Ratio	12.7%	9.1%	3.7%
	Sharpe Ratio	1.15	0.91	0.40

5.4. Discussion

Through the search for the optimal sentiment indicator based strategy, we find that the lag-1 news sentiment and lag-2 tweets sentiment are the most dominant factors in the formulation of the optimal trading strategy. In other words, the trading signals based on the sentiment feedback strength indicator rely significantly on business news articles published one day ago and tweet messages generated by the Twitter financial community two days ago. The key findings suggest that the combination of the two factors generates the best performance in terms of Sterling ratio and the percentage of winning trades. Furthermore, the lag-2 tweets sentiment exhibits a stronger effect on triggering trading signals over the lag-1 news sentiment, demonstrated by the higher parameter weight determined by the algorithm. On the contrary, the lag-1 news sentiment displays a greater sensitivity in affecting market returns, reflected by the lower summation threshold in the sentiment feedback strength indicator. In terms of trading frequency, the optimal sentiment indicator based strategy generates more trading signals at the latter portion of the evaluation period (see Figure 8). This is reflective of its adaptive feature to increasing market-wide volatility.

Another observation involves the chosen period of the study from 2012 to 2015, which signifies the recovery period from the 2008 financial crisis. Despite the strong bull market, our trading methodology has shown to provide significant out-performance relative to the benchmark in terms of risk-adjusted returns. In addition, the strategy fares well across different market conditions with the consideration of transaction costs. In Figure 8, the market is relatively stable in 2013 and 2014. Our algorithm trades less frequently, reflective of the longer holding period compared to the end of 2014 and the beginning of 2015. With more volatile market condition in 2015, the algorithm generates more trades and incurs higher transaction cost.

This study introduces a genetic programing approach to develop an optimal trading strategy with news and tweet sentiments. The proposed feedback strength indicator, a measure of the joint momentum between the news and tweet sentiments, was found to provide a significant improvement in trading performance over the S&P 500 financial market index ETF. Our analysis shows that the sentiment indicator not only yields higher returns over the evaluation period, but it also signals against substantial downside risk. In addition, the technical indicator strategy and buy-and-hold strategy yield less desirable results with higher volatility and lower returns. The evidence presented in this study highlights the value of the feedback strength indicator using news sentiment and tweet sentiment, and further demonstrates that a sentiment based trading strategy can be constructed to exploit market anomalies caused by major sentiment momentum spikes.

6. Conclusion

This study presents a novel framework for developing a sentiment feedback strength based trading strategy using genetic programming. Motivated by the empirical phenomenon that news and social media exhibit persistent and predictive power on financial market movement, we propose a sentiment indicator based on feedback strength between the news and tweet sentiments. By quantifying the joint momentum of the sentiment series, we can detect significant market anomalies that can be exploited for a significant improvement on trading performance. We find that the sentiment indicator based genetic programming approach yields superior market returns with low average monthly maximum drawdown over the period from 2012 to 2015. When comparing the Sterling ratio and other risk measures, the proposed sentiment indicator based strategies are superior to the technical indicators and the traditional buy-and-hold strategy. The out-performance suggests that news and tweet sentiments can be regarded as valuable sources of information in constructing meaningful trading system along with technical indicators.

For future work, we aim to explore the trading performance of the sentiment feedback strength based strategy using intraday data. With finer time scales, the signals based on the joint momentum of the sentiment series may lead to more profitable short-term trading opportunities. The other area of interest is to detect events with abnormal sentiment spikes and to investigate their effects on the financial markets.

References

[1] M. Baker, J. Wurgler, The Journal of Economic Perspectives (2007) 129–151.

[2] N. Barberis, A. Shleifer, R. Vishny, Journal of Financial Economics 49 (1998) 307–343.

[3] K. Daniel, D. Hirshleifer, A. Subrahmanyam, the Journal of Finance 53 (1998) 1839–1885.

[4] K. D. Daniel, D. Hirshleifer, A. Subrahmanyam, The Journal of Finance 56 (2001) 921–965.

[5] N. Chopra, J. Lakonishok, J. R. Ritter, Journal of financial Economics 31 (1992) 235–268.

[6] R. L. Porta, J. Lakonishok, A. Shleifer, R. Vishny, The Journal of Finance 52 (1997) 859–874.

[7] A. Shleifer, Inefficient markets: An introduction to behavioral finance, Oxford university press, 2000.

[8] P. C. Tetlock, The Journal of Finance 62 (2007) 1139–1168.

[9] P. C. Tetlock, M. Saar-Tsechansky, S. MacSkassy, Journal of Finance 63 (2008) 1437–1467.

[10] J. E. Engelberg, A. V. Reed, M. C. Ringgenberg, Journal of Financial Economics 105 (2012) 260–278.

[11] M. Baker, J. Wurgler, Y. Yuan, Journal of Financial Economics 104 (2012) 272–287.

[12] G. W. Brown, M. T. Cliff, Journal of Business (2005) 405–440.

[13] D. García, Journal of Finance 68 (2013) 1267–1300.

[14] A. Kurov, Journal of Banking & Finance 34 (2010) 139–149.

[15] B. M. Barber, T.-H. Ho, T. Odean, Management Science 58 (2012) 1–1.

[16] J. Bollen, H. Mao, X. Zeng, Journal of Computational Science 2 (2011) 1–8.

[17] X. Zhang, H. Fuehres, P. A. Gloor, Procedia - Social and Behavioral Sciences 26 (2011) 55 – 62. URL: http://www.sciencedirect.com/science/article/pii/S1877042811023895. doi:http://dx.doi.org/10.1016/j.sbspro.2011.10.562, the 2nd Collaborative Innovation Networks Conference - {COINs2010}.

[18] S. Y. Yang, S. Y. K. Mo, A. Liu, Quantitative Finance 15 (2015) 1637–1656.

[19] Q. Li, T. Wang, Q. Gong, Y. Chen, Z. Lin, S.-k. Song, Decision Support Systems 61 (2014) 93–105. URL: http://linkinghub.elsevier.com/retrieve/pii/S0167923614000232. doi:10.1016/j.dss.2014.01.013.

[20] M. Piškorec, N. Antulov-Fantulin, P. K. Novak, I. Mozetič, M. Grčar, I. Vodenska, T. Smuc, Scientific reports 4 (2014).

[21] B. Johnson, Algorithmic Trading and DMA: An introduction to direct access trading strategies, 4Myeloma Press, London, 2010.

[22] L. A. Smales, Finance Research Letters 11 (2014) 122–130.

[23] L. A. Smales, Journal of Banking & Finance 49 (2014) 275–286.

[24] S. Y. K. Mo, A. Liu, S. Yang, SSRN Working Paper (2015).

[25] M. Dempster, C. Jones, Quantitative Finance 1 (2001) 397–413.

[26] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, D. C. L. Ngo, Expert Systems with Applications 42 (2015) 306–324. URL: http://linkinghub.elsevier.com/retrieve/pii/S0957417414004801. doi:10.1016/j.eswa.2014.08.004.

[27] N. J. Ferguson, D. Philip, H. Y. T. Lam, J. M. Guo, Multinational Finance Jounal 19 (2015) 1–31.

[28] J.-S. Chen, C.-L. Chang, J.-L. Hou, Y.-T. Lin, Expert Systems with Applications 35 (2008) 273–278.

[29] L. Mitra, G. Mitra, D. Di Bartolomeo, Quantitative Finance 9 (2009) 887–895.

[30] A. Healy, A. W. Lo, Managing real-time risks and returns: The Thomson Reuters NewsScope event indicies, John Wiley & Sons, 2011.

[31] D. Leinweber, J. Sisk, Relating News Analytics to Stock Returns, The Handbook of News Analytics in Finance, 2012.

[32] L. K. C. Chan, J. Lakonishok, The Journal of Finance 52 (1997) 713–735.

[33] D. B. Keim, A. Madhavan, Financial Analysts Journal 54 (1998) 50–69.

[34] J. A. Bikker, L. Spierdijk, P. J. van der Sluis, Journal of International Money and Finance 26 (2007) 974–1000.

[35] S. Y. Yang, S. Y. K. Mo, X. Zhu, in: Computational Intelligence for Financial Engineering & Economics (CIFEr), 2104 IEEE Conference on, pp. 55–62.

[36] S. Baccianella, A. Esuli, F. Sebastiani, in: Language Resources and Evaluation, volume 10, pp. 2200–2204.

[37] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Y. Wah, D. C. L. Ngo, Expert Systems with Applications 42 (2014) 306–324.

[38] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Computational linguistics 37 (2011) 267–307.

[39] A.-D. Vo, C.-Y. Ock, in: Computational Collective Intelligence. Technologies and Applications, Springer, 2012, pp. 373–382.

[40] A. Muhammad, N. Wiratunga, R. Lothian, R. Glassey, in: Proceedings of the BCS SGAI: Workshop on Social Media Analysis 2013, pp. 7–18.

[41] A. W. Lo, H. Mamaysky, J. Wang, The Journal of Finance 55 (2000) 1705–1770.

[42] J. Fang, Y. Qin, B. Jacobsen, Journal of Behavioral and Experimental Finance 4 (2014) 25–56.

[43] X. Zhang, Y. Hu, K. Xie, W. Zhang, L. Su, M. Liu, Knowledge-Based Systems (2014). URL: `http://linkinghub.elsevier.com/retrieve/pii/S0950705114002925`. doi:`10.1016/j.knosys.2014.08.010`.

[44] G. Appel, The moving average convergence-divergence trading method: advanced version, Scientific Investment Systems, 1985.

[45] T. Aspray, Technical Analysis of Stocks & Commodities 7 (1989) p56–61.

[46] J. W. Wilder, New concepts in technical trading systems, Trend Research Greensboro, NC, 1978.

[47] J. R. Koza, Genetic programming: on the programming of computers by means of natural selection, volume 1, MIT press, 1992.

[48] J. H. Holland, Scientific American 267 (1992) 66–72.

[49] W. Banzhaf, P. Nordin, R. E. Keller, F. D. Francone, Genetic programming: an introduction, volume 1, Morgan Kaufmann San Francisco, 1998.

[50] T. Preis, H. S. Moat, H. E. Stanley, Scientific Reports 3 (2013) 1684.

[51] D. Sun, Y. Du, W. Xu, M. Zuo, C. Zhang, J. Zhou, Pacific Asia Journal of . . . 6 (2015) 19–37.

[52] H. Bessembinder, K. Chan, Pacific-Basin Finance Journal 3 (1995) 257–284.

[53] W. M. Fong, L. H. M. Yong, Journal of Empirical Finance 12 (2005) 43–76.

[54] A. Ito, Pacific-Basin Finance Journal 7 (1999) 283–330.

[55] R. D. Harris, F. Yilmaz, Journal of Banking and Finance 33 (2009) 1575–1585.

Appendix

Table 7: Tweets sentiment and returns empirical relation.

Number of Lags	Tweets→Returns (Coefficient)	Tweets→Returns (p-value)	Returns→Tweets (Coefficient)	Returns→Tweets (p-value)
1	0.1121	0.028**	0.0417	0.865
2	0.0810	0.114	0.2367	0.328
3	-0.0137	0.791	-0.1523	0.533
4	-0.0551	0.292	0.0083	0.973
5	0.0162	0.758	0.1374	0.577

Note: *** 1% confidence level,** 5% confidence level, and * 10% confidence level.

Table 8: News sentiment and returns empirical relation

Number of Lags	News→Returns (Coefficient)	News→Returns (p-value)	Returns→News (Coefficient)	Returns→News (p-value)
1	-0.0012	0.802	2.4097	2.71e-11***
2	-0.0046	0.334	0.8352	0.024**
3	-0.0066	0.165	0.7226	0.050
4	-0.0117	0.014**	0.1135	0.760
5	-0.0079	0.097	0.1477	0.692

Note: *** 1% confidence level,** 5% confidence level, and * 10% confidence level.

Table 9: News sentiment and Tweets sentiment empirical relation

Number of Lags	News→Tweets (Coefficient)	News→Tweets (p-value)	Tweets→News (Coefficient)	Tweets→News (p-value)
0	-0.0042	0.62	-0.0933	0.62
1	-0.0004	0.96	-0.2644	0.16
2	-0.0136	0.10	-0.1663	0.37
3	-0.0175	0.04 **	-0.2460	0.19
4	-0.0116	0.17	-0.1969	0.29
5	-0.0085	0.31	-0.0096	0.96

Note: *** 1% confidence level,** 5% confidence level, and * 10% confidence level.