S-Factors™: Definition, Use, and Significance

Social Market Analytics, Inc.

Harness the Power of Social Media Intelligence

January 2014

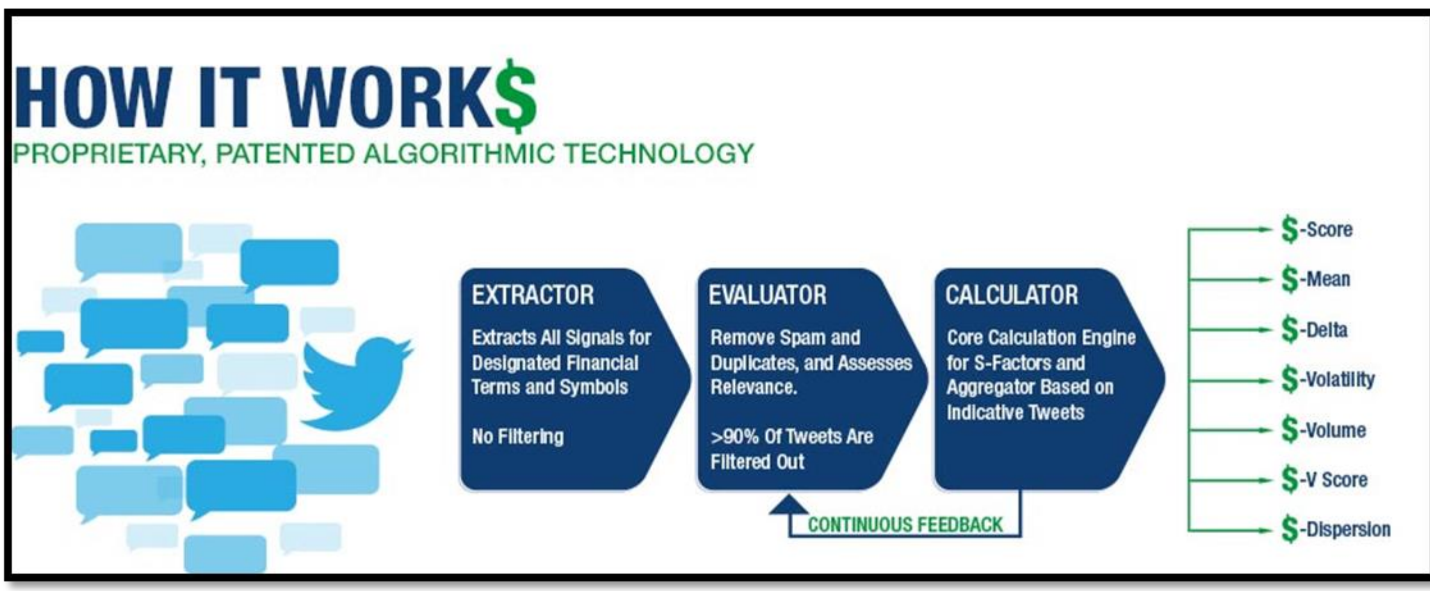
Introduction

Social Market Analytics, Inc., (SMA) produces a family of metrics, called **S-Factors™**, designed to capture the signature of financial market sentiment. SMA applies these metrics to social media data to estimate sentiment for indices, sectors, and individual securities, yielding and recording time series of these measurements on daily and intraday time scales.

The Definition of S-Factors™

SMA computes **S-Factors™** on all U.S. securities that have sufficient social media data. In general, positive S-Score™ levels are associated with favorable market sentiment, while negative levels are unfavorable. We expect changes in market sentiment, as measured by changes in S-Score™, and associated metrics, to be reflected in stock price changes over some time horizon.

SMA's servers poll Twitter's API with access to over 400 million daily tweets and sift out stock-specific tweets observed during a time sample window. The collected tweets are filtered for financial trading relevance and scored for market sentiment content. Then, the tweet scores are aggregated to produce a sentiment measurement, at an observation time, for each stock implied from data observed during the sample window.



S-Factors™ are designed to quantify sentiment for stocks, market sectors, and industries. **S-Factors™** provide a perspective to understand changes in sentiment levels and reveal the signature of market sentiment over time.

SMA's processing engine delivers these **S-Factors™** metrics:

S-Score™: is the normalized representation of a sentiment time series over a lookback period. S-Score™ is a measure of the deviation of a stock's sentiment intensity level from a normal state. S-Score™ levels greater than 2.0 or less than -2.0 indicate significant positive or negative sentiment states.

S-Volume™: is the volume of indicative tweets contributing to a sentiment estimate at an observation time. A significant change in S-Volume™ over time is a good indicator of unusual social media commentary on a stock.

S-Dispersion™: is a measure of the diversity of Twitter sources contributing to a sentiment estimate at an observation time. Dispersion levels range from 0.0 to 1.0. A level of 1.0 indicates that all indicative tweets captured for a stock come from distinct Twitter accounts, while small dispersion levels, approaching 0.0, indicate that a small number of sources produced commentary on the stock.

S-V Score™: is the normalized representation of a stock's indicative tweet volume time series over a look back period and is a measure of the deviation of a stock's indicative tweet volume level from a normal state.

S-Mean™: is the average level of a stock's sentiment time series over a look back period.

S-Volatility™: is a measure of the variability of a stock's sentiment time series over a look back period**.**

S-Delta™: is the change in S-Score™ level at an observation time relative to an earlier time, and is a first order measuremet of a stock's sentiment trend.

SMA employs a three stage processing pipeline to mine **S-Factors™** from the Twitter message stream. This process is performed 24/7 for all constituents of the SMA universe yielding estimates at regular observation times throughout each day. Each component of the processing pipeline is described as follows.

EXTRACTOR

The Extractor accesses the API web services of Twitter and GNIP, a microblogging data aggregator, to captures groups of tweets. A Data Acquisition process polls these sources to capture tweets containing commentary on the members of the SMA stock universe. The polling process continuously cycles through the universe list, adaptively polling for list members with current, active content in the message stream. Data Acquisition receives tweets as JSON document objects and retains these raw tweet data for real-time and historical analysis. Metadata Extraction and Reference Data Extraction processes further mine features of the tweets (language, location, posting time) and features of the Twitter user account that originated the tweets (user profile, followers, account rating).

EVALUATOR

The Evaluator analyzes each tweet and develops a list of tweets with financial market relevance to the entities in the SMA stock universe. These are called "indicative" tweets, as these indicate expressions of market trading sentiment for these stocks. The sentiment scores of indicative tweets contribute to an entity's sentiment signature. The Evaluator uses established Natural Language Processing parsing tools, which SMA has enhanced and tuned for performance in the domain of financial markets. These tools perform tokenization of tweet text to identify words, phrases, and stock symbols in the captured tweets. The tokenized tweets are passed to a Duplicate Detection process, which works to eliminate duplicate tweets submitted from the same source and to identify the occurrence and sources of "re-tweets". Duplicate Detection and re-tweet policies work to reduce the influence of tweets originating from "spamming" users, reducing the noise level of the tweet stream and improving signature estimates.

A Subject Mention Detection process identifies specific entities contained in the text of the tweets. Tweets having content on members of the sentiment Universe are labeled as "relevant" and are tagged for the corresponding entities. A Relevance Assessment process is the final step at the Evaluator stage and proceeds to analyze the set of relevant tweets with respect to ratings, developed by SMA, for the Twitter accounts that are the originators of the captured tweets. Operational experience has shown that 90% of the tweets, determined to be relevant, originate from about 10% of the observed Twitter accounts. The 10% portion of accounts that present relevant content are also high volume accounts, meaning that these accounts have much higher total number and frequency of tweets presented over time compared to other accounts. Further, accounts originating tweets that have content determined to be NOT relevant are typically from low volume, sporadically active Twitter users. A filter is applied to eliminate tweets from low volume originators transforming the set of relevant tweets to the set of indicative tweets for the entities that are used in the signature estimation stage.

CALCULATOR

The Calculator determines the sentiment signatures for each member of the SMA stock universe. For each entity, a set of indicative tweets is received by a Sentiment Calculation process and accesses the tokenized results for each tweet. The sentiment level for each word parsed from a tweet is obtained from a domain specific Sentiment Dictionary. SMA uses a Sentiment Dictionary tuned for performance in the financial market domain. Currently, SMA's Sentiment Dictionary has about 18,000 words (uni-grams) and 400 two word phrases (bi-grams) that have content and sentiment levels of relevance to financial market activity as expressed in microblogging messages. The sentiment score for a tweet is the average value of the total sentiment content identified from tokenization. The sentiment level of each tweet is mapped onto a fine grained, continuous range from -1.0 (most negative) to +1.0 (most positive).

At any time, the raw sentiment level inferred from associated tweets is the simple aggregate of all indicative tweet sentiment levels captured during the prior 24 hours for each entity. A Bucketing and Weighting process operates on an entity's indicative tweets and groups these into time period buckets based on the arrival time of each tweet. A Normalization and Scoring process calculates the S-Score™ and other **S-Factors™** for each entity with active content at the time of the estimate. An Aggregation Scoring process calculates **S-Factors™**

and market indices, market sectors and industry groups by aggregating the S-Score™ of the constituent entities. These processes are performed for all members of the SMA stock universe and yield time series of **S-Factors™** signatures for each entity.

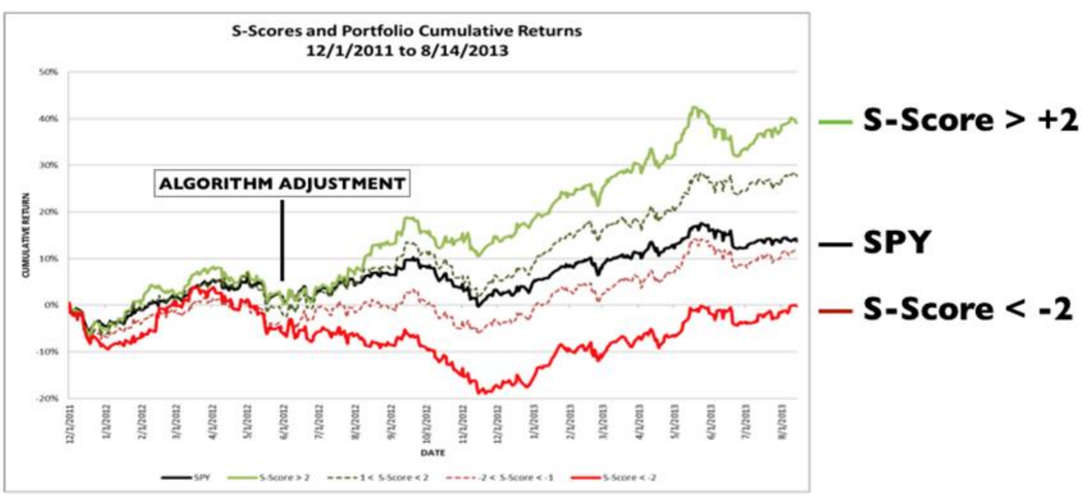
Interpretation of S-Factor™ Behavior

Relationship of Pre-market Sentiment Levels to Stock Price Changes

The most fundamental application of **S-Factors™** to market sentiment analytics is to consider the behavior of the level of the S-Score™ metric with respect to market returns. In general, positive S-Scores™ are associated with favorable changes in investor sentiment, while negative levels are associated with unfavorable changes. We expect investor sentiment changes to result in stock price changes. Similarly, we expect larger changes in investor sentiment to be associated with larger stock price changes.

S-Score™	Market Sentiment Regime
> 3	Extreme Positive
> 2 and < 3	High Positive
> 1 and < 2	Positive
> -1 and < 1	Neutral
< -1 and > -2	Negative
< -2 and > -3	High Negative
< -3	Extreme Negative

Historical Cumulative Returns



From December 1, 2011 through August 14, 2013, cumulative returns calcuations illustrates the power of S-Scores™. A hypothetical sequence of portfolios constructed from stocks with S-Scores™ of 2 or higher delivered cumulative returns of 39.05% (green line) compared to the S&P 500 at 13.75%. Stocks satisfying this criterion are assumed to be purchased at the Open and sold on the Close each day.

Portfolios constructed from stocks with S-Scores™ of -2 or lower significantly underperformed the S&P 500 (**red line**).

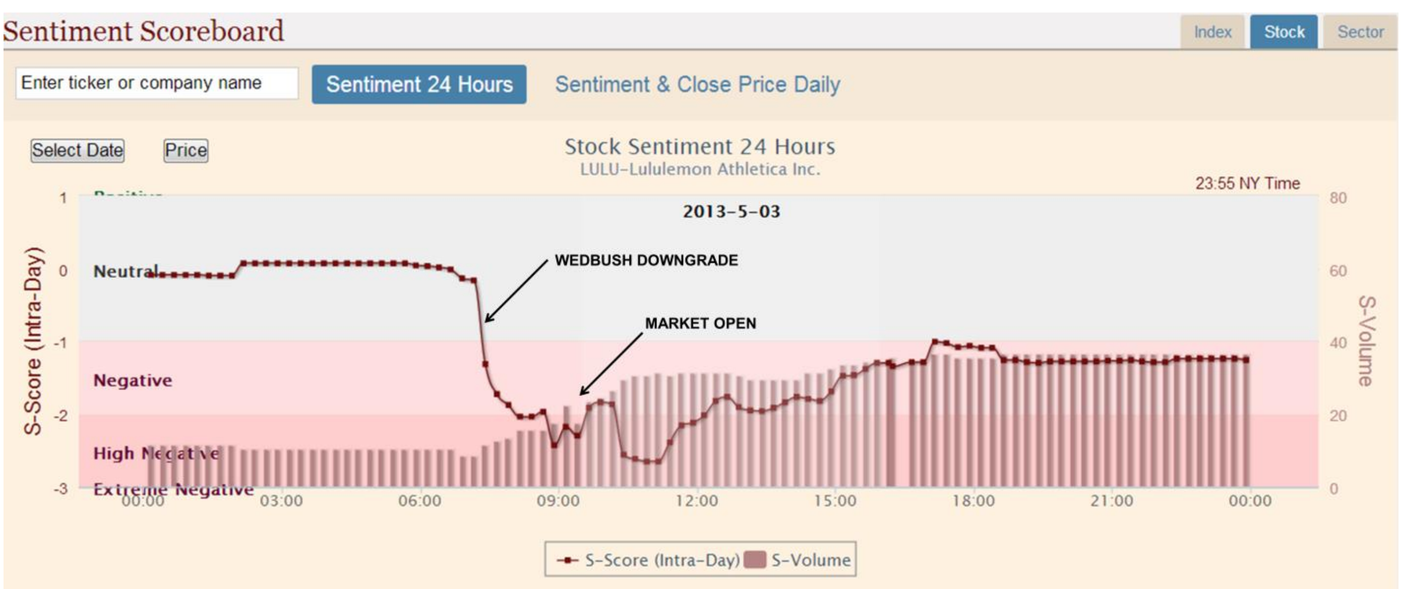
For stocks with positive S-Scores™, the cumulative returns realized on this interval produced risk adjusted returns with outstanding Sharpe Ratios.

Buckets	Return	Sharpe Ratio
S-Score™ > 2	39.05%	1.72
1< S-Score™ < 2	27.68%	1.31
Universe	18.11%	0.93
SP 500	13.75%	0.77
-2< S-Score™ > -1	11.38%	0.61
S-Score™ > -2	-0.1%	-0.06

CASE STUDIES

Here are some examples of **S-Factors ™** in practice.

Lululemon (LULU) May 3, 2013

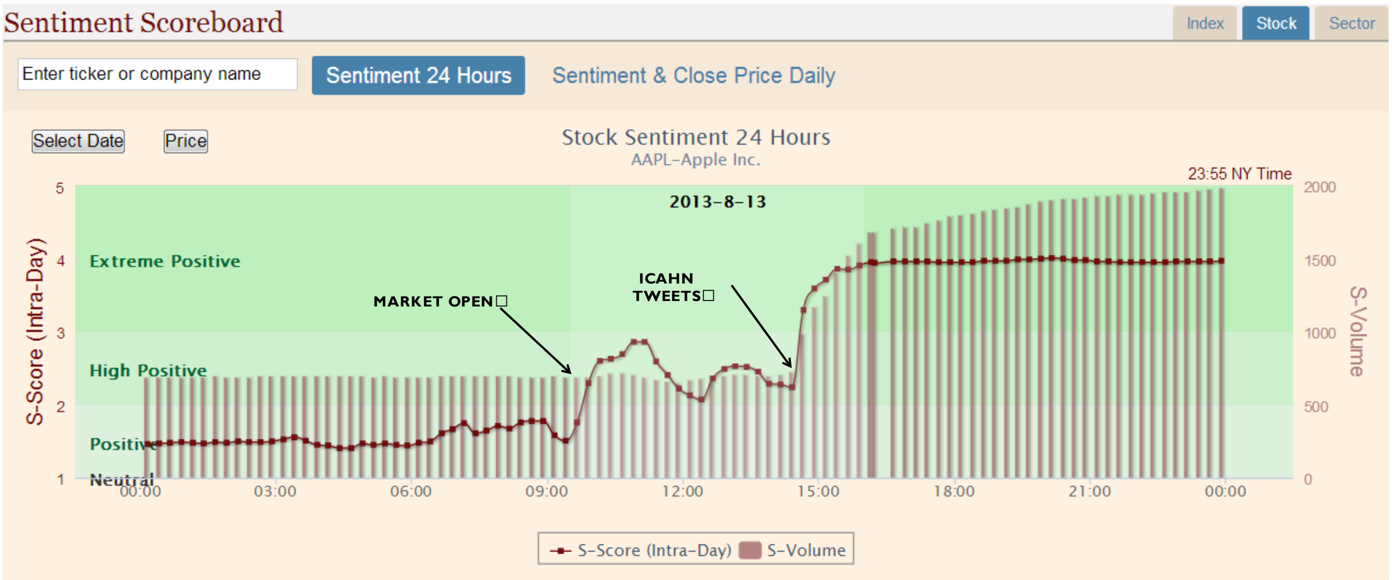


About 2 hours before the market open of May 3, 2013, a notable analyst at Wedbush, Inc. issued a downgrade on the stock. At that time, LULU's S-score decreased rapidly from neutral to negative, reaching high negative levels by the market open. The decline in S-score was coincident with a steady increase in LULU's S-volume indicating increased commentary on the stock. The stock price did not react so dramatically, opening trading on May 3 at $76.05, a small gap down from the previous day's close. LULU traded lower during the day, but rallied to close at $76.36, a loss of about 1% on the day. Market reaction to the downgrade developed over the weekend of May 4 and 5. On May 6, LULU opened slightly gap down at $76.51 and proceeded to sell off during the day to close at $74.40, a decline of 9.8% from the open of the May 3 session.

Whole Foods Market (WFM) May 7, 2013

On May 7, 2013, Whole Foods Market, Inc., reported earnings at 4 pm, after the market close. WFM's earnings beat estimates and the stock price surged in after hours trading. During the market session of May 7, WFM's S-score increased steadily throughout the day. In particular, during the last hour of trading (after 3 pm Eastern), WFM's S-volume increased rapidly, coincident with a run up in the stock price into the close. WFM closed the May 7 session at $92.80, up 1% on the day. On May 8, WFM opened at $100.70, a gap up of 8.5% from the previous close. WFM continued to rally to end at $102.19, a gain of 10.1% on the day. A long position, entered any time after 1 pm Eastern on May 7 would have realized a good gain if closed out any time on May 8.

Apple (AAPL) August 13, 2013



On August, 13, 2013, Apple Inc, presented a striking example demonstrating the behavior of social media sentiment surrounding events that influence stock prices. At 2:21 pm EDT, outspoken billionaire investor, Carl Ichan, announced via Twitter that his firm had taken a "large position" in AAPL. The signature of this event quickly emerged and was detected by SMA's processing. AAPL's S-Score and S-Volume metrics surged rapidly as the event propagated through Twitter. AAPL's stock price experienced a run up into the close to end the market session at $489.57, a gain of 4.8% on the day. AAPL opened at $470.94 on August 13, but increased steadily through the morning reaching $475.20 by 2 pm EDT. This price action was coincident with sustained High Positive S-Score levels, prior to Icahn's announcement, and offered an opportunity to enter a long position at very favorable prices. As traditional news services initiated coverage of the event after 2:21 pm, AAPL's S-Score moved solidly to Extreme Positive levels, which continued into August 14, as the stock gapped up at the open to $497.88. The stock rallied to $504.25 by 12:45 pm then entered a mild sell off as profit taking ensued. AAPL ended the day at $498.50, up 5.8% from the open of August 13.