S-Factors™: Their Definition, Use, and Significance

White Paper – SMA-04

Social Market Analytics, Inc.

Actionable Intelligence from Social Media Data for Financial Markets

August 2015

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 data captured from social media sources to estimate sentiment for financial instruments, yielding and recording time series of these measurements on a variety of intraday time scales. Our servers process social media data streams to produce sentiment estimates for all active equities, major indices, sectors, industries, commodities and currencies and ETFs.

This note provides an introduction to Sentiment Factors (or S-Factors™), their interpretation and use, and is presented in three sections. First, the definition of these metrics is reviewed. Next, the behavior of these metrics over time and their interpretation is discussed. The note concludes with a section listing concepts for the application of S-Factor™ metrics.

The Definition of S-Factors™

SMA computes S-Factors™ for equities, commodities, currencies, indices and ETFs. 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 price changes.

For each entity in SMA's universe, our servers poll the Twitter and StockTwits APIs to capture Tweets about the entity 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 the entity implied from data observed during the sample window.

SMA employs a three stage processing pipeline to mine **S-Factors™** from the Twitter and StockTwits message streams. This process is performed 24/7 for all constituents of the SMA universe (covering ~ 3700 stocks) yielding estimates at regular observation times throughout each day. Figure below reviews these processing stages with further detail available in our patent ["Systems and Methods of Detecting, Measuring, and Extracting Signatures of Signals Embedded in Social Media Data Streams", U.S. Patent No 9,104,734, August 2015, Washington, DC: U.S.].

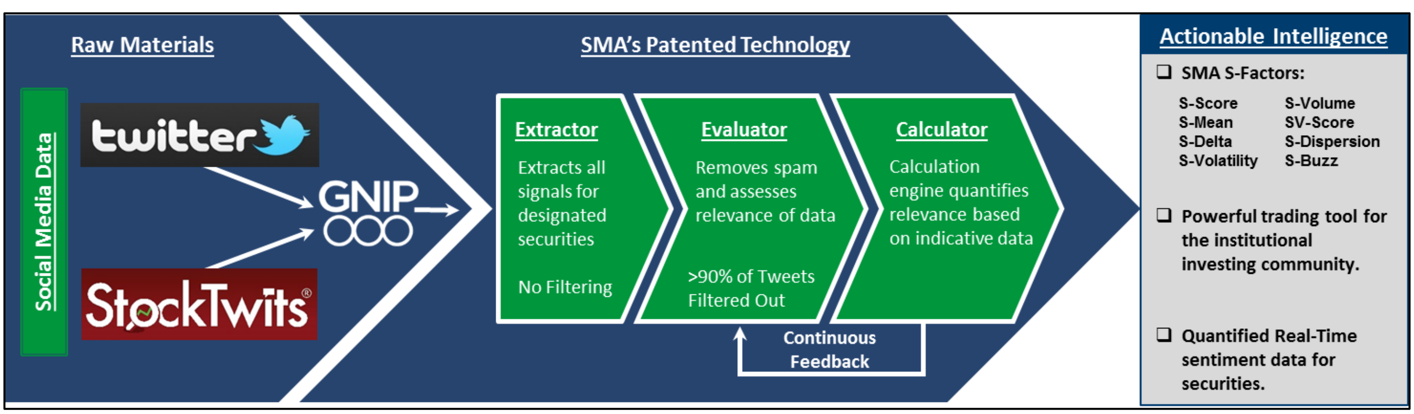


Figure 1: Social Market Analytics Processing Architecture

Extractor - The Extractor accesses the API web services of Twitter and GNIP, a micro-blogging data aggregator. 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 securities with current content in the message stream.

Evaluator - The Evaluator analyzes each Tweet for 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 Evaluator uses proprietary natural language processing algorithms to asses each message. SMA also uses a proprietary algorithm to further filter specific twitter accounts. In short, SMA utilizes Tweet content and characteristics of the individuals Tweeting to distill the intentions of professional investors.

Calculator - The Calculator determines the sentiment signatures for each member of the SMA stock universe. 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.

S-Factors™ are designed to quantify sentiment for stocks, commodities, currencies, Indices, market sectors, and industries. S-Factors™ provide a perspective for understanding changes in sentiment levels and reveal the signature of market sentiment over time. A visual representation of the S-Factors is as shown below.

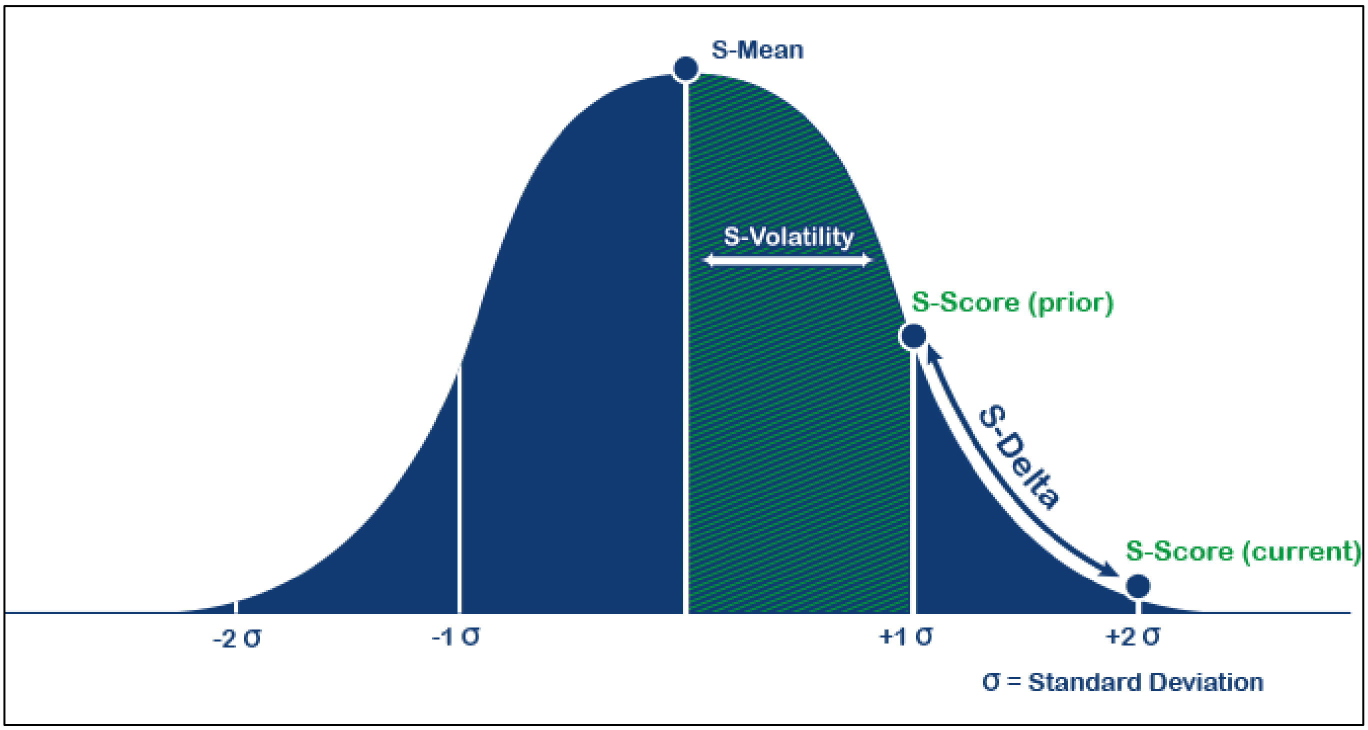


Figure 2: Visual Representation of S-Factors™

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

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-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-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.

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

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-Buzz™: is a measure of abnormal volume activity, which compares a stock's sentiment volume level to the average volume level measured for the sentiment Universe at an observation time. Here, the term "indicative" means Tweets that pass SMA's filtering processes and are used in sentiment estimates.

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

S-Poster™: is the number of unique posting accounts that have contributed to indicative Tweets in calculating the sentiment estimate. A high number of S-Poster or a significant change in is a good indication of increase in breadth of the commentary on a stock

SP-Score™: is the normalized representation of unique accounts with indicative Tweets for a stock in time series over a look back period. SP-Score™ is a measure of the deviation of number of unique accounts Tweeting for a stock from a normal state.

Predictive analytics applied to social media and financial markets is a new, rapidly evolving technology, offering many opportunities for innovation. SMA is committed to leading-edge development of social media analytics for the benefit of its product line and client base.

A complete list of all the sentiment metrics is displayed below.

Factor Name	Description
Raw-S	Unweighted Sentiment Estimate
Raw-S-Mean	20 Day Moving Average of Raw-S
Raw-Volatility	20 Day Moving Standard Deviation of Raw-S
Raw-Score	Normalized Value of Raw-S
S	Exponentially Weighted Sentiment Estimate
S-Mean	20 Day Moving Average of S
S-Volatility	20 Day Moving Standard Deviation of S
S-Score	Normalized Value of S. This Is SMA's S-Score
S-Volume	Indicative Tweet Volume Used To Compute The Sentiment Estimate
SV-Mean	20 Day Moving Average of S-Volume
SV-Volatility	20 Day Moving Standard Deviation of S-Volume
SV-Score	Normalized Value of S-Volume
S-Dispersion	Measurement of The Tweet Source Diversity Contributing To A Sentiment Estimate
S-Buzz	Measurement of Unusual Volume Activity
S-Delta	Change In S-Score Over A Look back Period

Interpretation of S-Factor™ Behavior

Relationship of Pre-market Sentiment Levels to Stock Price Change

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. Table 1 shows the connection between S-Score™ level and sentiment regimes.

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

Table 1: Sentiment Regimes

To investigate whether pre-market-open sentiment measurements are predictors of post-market-open price change, we created a scatter plot for overnight (C-O) and open-to-close (O-C) price changes. Each point in the scatter plot corresponds to the centroid for individual sectors which had securities with large S-Scores™. The plot also includes the centroid for securities with large S-Scores and the SPY. For the purpose of this note, stocks with S-Scores™ > 2 and S-Scores™ < -2 were selected.

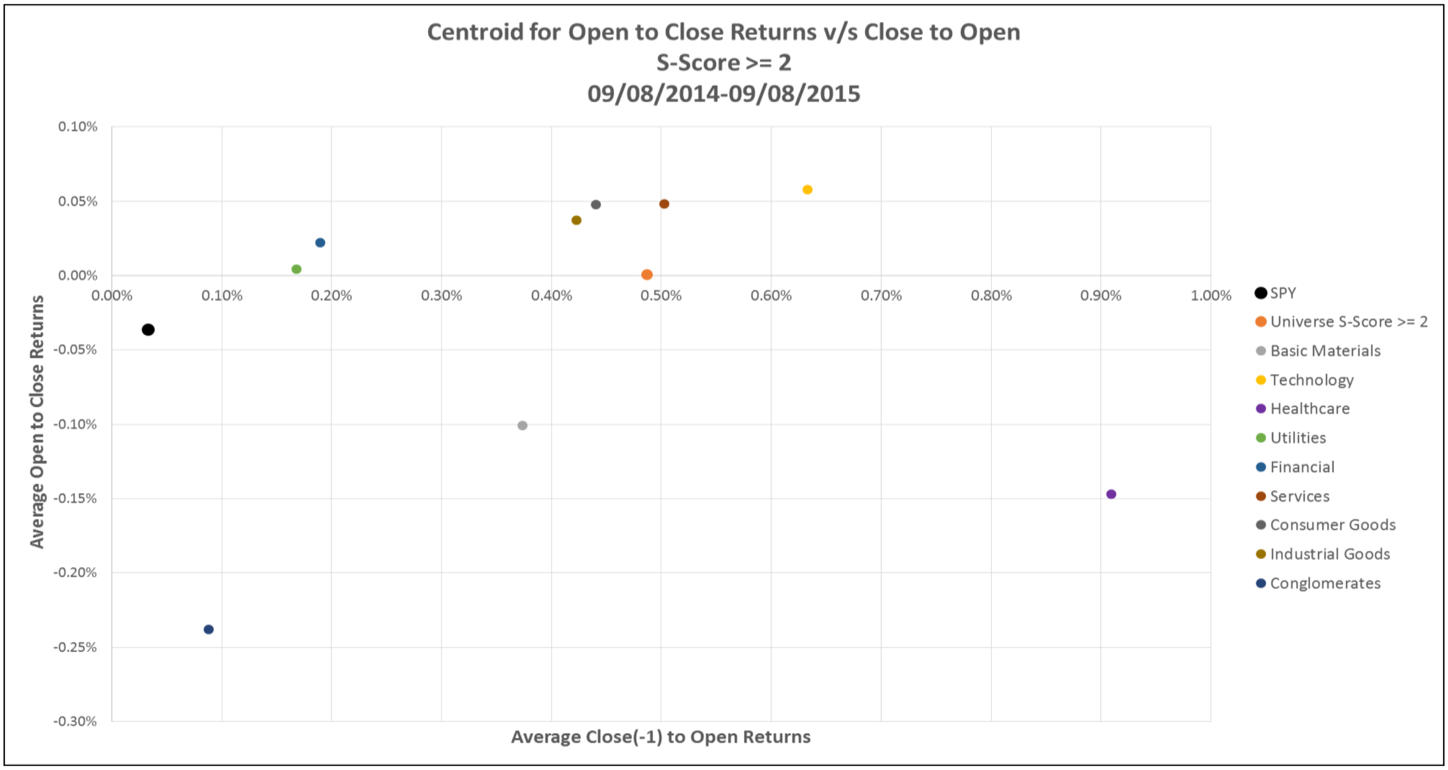


Figure 3: Overnight and Open-To-Close Price Returns, Positive Sentiment.

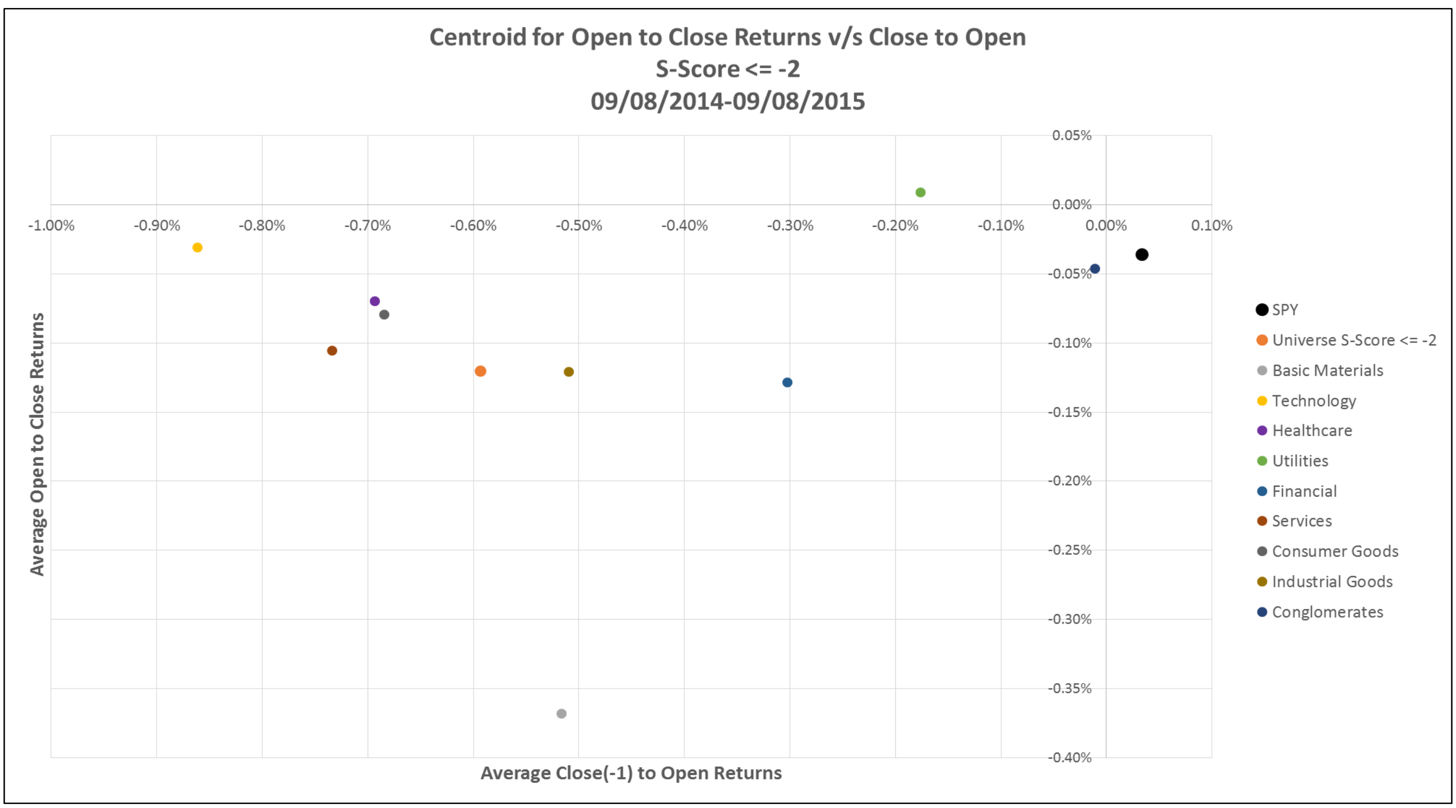


Figure 4: Overnight and Open-To-Close Price Returns, Negative Sentiment.

The horizontal axis is the percentage overnight price change (C-O). The vertical axis is the percentage open-to-close price change (O-C). These data are from a sample space of events for the SMA universe observed during the last year. The plots show the events that satisfy the indicated conditions on S-Score™.

The scatter plots show that pre-market-open S-Score™ is a very good estimator of the direction of market-at-open price changes for most of the sectors. The scatter plot of S-Score™ > 2 has most sector points in the upper quadrant while the scatter plot with S-Score™ < -2 has most sector points in the lower quadrant. The chart also also illustrates that the universe of stocks with S-Score magnitude over 2 has significant outperformance on the positive side and underperformance on the short side as compared to the SPY benchmark. This simple observation suggests that SMA has accurate sentiment scoring algorithms.

Below is a summary of observed open-to-close returns. The relatively large frequency of moves of greater than + 2% and less than -2%, demonstrates the distribution has fat tails, and is evidence that S-Score™ is not normally distributed.

S-Score > 2	Events	Percentage
Total	49487	100.00%
Up	23831	48.16%
> 1%	12894	26.06%
> 2%	7123	14.39%

S-Score < -2	Events	Percentage
Total	20427	100.00%
Down	10179	49.83%
< -1%	5826	28.52%
< -2%	3395	16.62%

Table 2: Relative Frequency of Overnight Returns.

For S-Scores™ > 2.5, 52% of open-to-close returns are up, 36.5% are up greater than 1% and 24% are up greater than 2%. For S-Scores™ < -2.5, 46.3% of open-to-close returns are down, 30% are down more than -1%, and 18.5% are down greater than -2%.

Merger and Acquisition Event: Announcement that TEVA will acquire AGN

Teva Pharmaceutical Industries surged in pre-market trading on July 27, 2015 on news that the company will be acquiring Allergan's (AGN) generic drug business. But before this happened, sentiment on Twitter had already become strongly positive. At 4:00 AM EDT, when the stock price was $66.00, there was significant positive sentiment on Twitter. The sentiment rapidly shifted positive in a matter of minutes. By 7:24 AM, the stock was trading at $72.30. The stock opened at $67.80 when the sentiment was 3.92 and closed at $72.

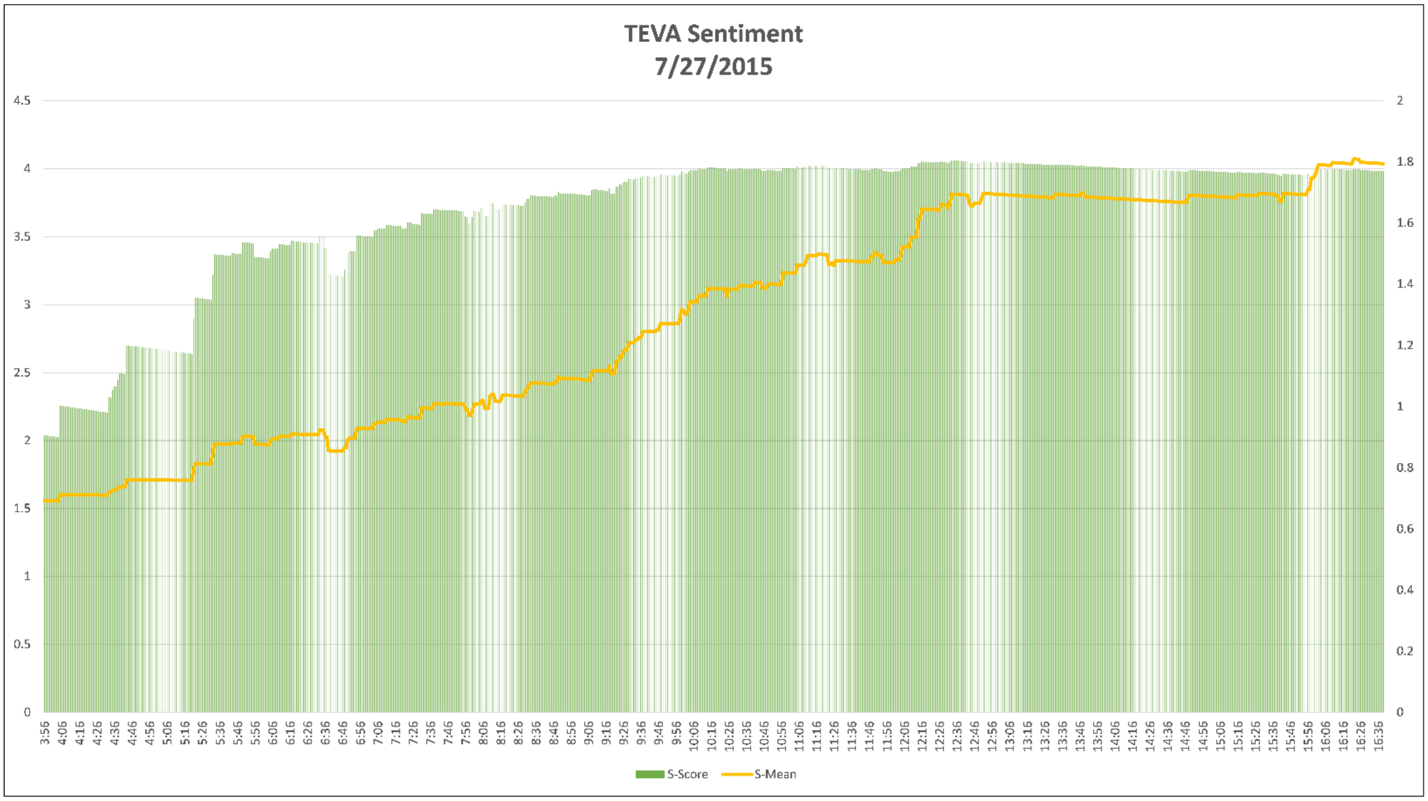
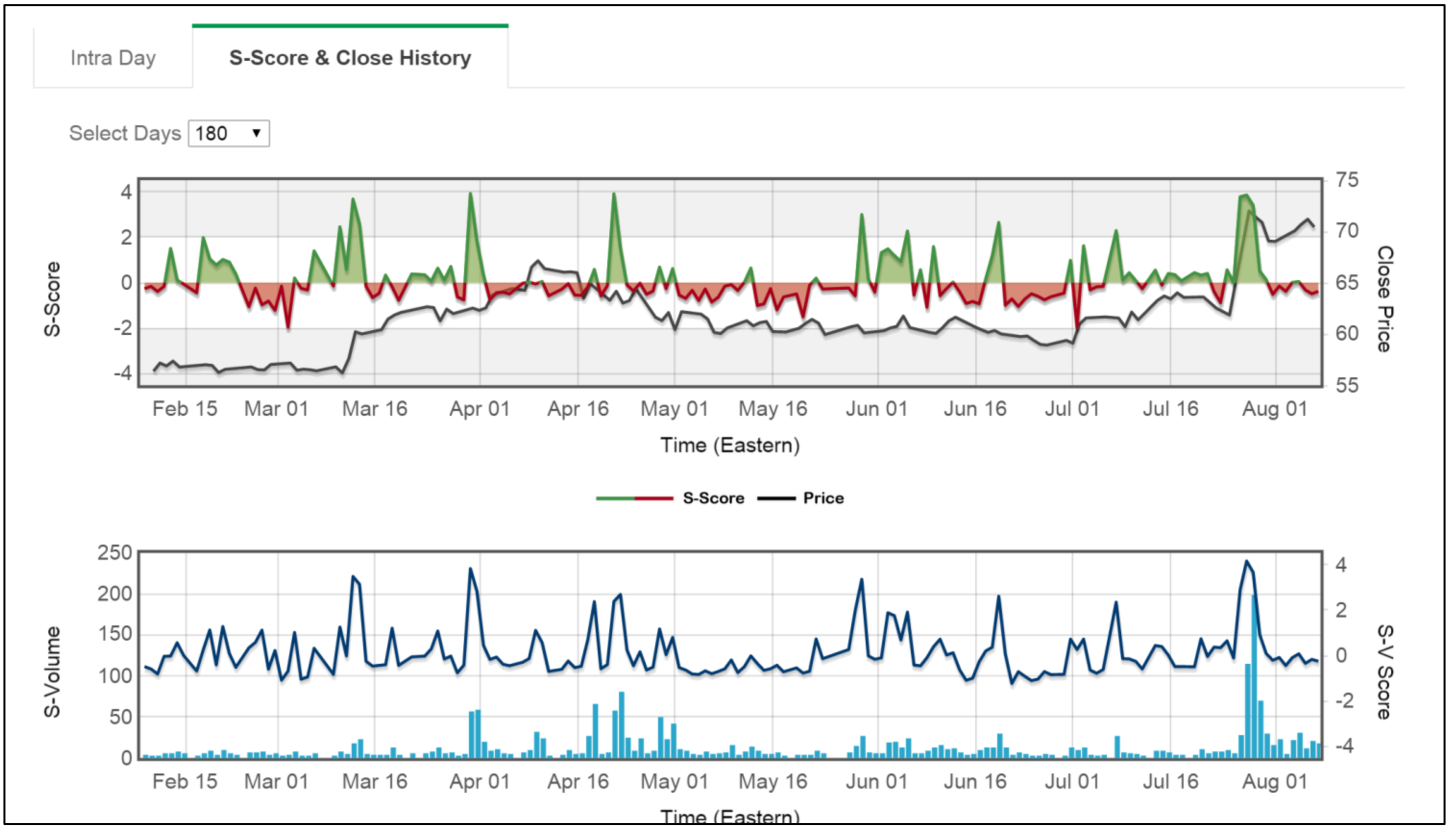


Figure 5: S-Score™ For TEVA Pre- and Post-Announcement.



Historically, daily sentiment scores for TEVA fluctuated near 0 (Neutral), with low social media activity as indicated by the time series of the S-Volume™ metric. This behavior started to change on July 26 with significant upticks in indicative Tweet volumes and sentiment levels. On the morning of July 277th, TEVA's S-Score™ increased sharply to a high positive level, coincident with a spike in S-Volume™ consistent with high social media activity, indicating that SMA's processing technology had successfully detected the signature of positive sentiment for TEVA embedded in the Twitter data stream. This high positive sentiment level persisted through the open on July 28th and then started to return to typical historical levels as the markets and social media fully integrated the effect of the announcement.

Rumored Announcement of Acquisition: Twitter (TWTR)

On July 14, 2015 at 11.39 AM EDT, a rumor started spreading on Twitter about Twitter's being bought by Bloomberg.

At 11:40 AM, there was a Tweet from user 'beckyhiu' indicating that Bloomberg had offered $31 Billion to buy Twitter and that Twitter was considering the offer. This rumor caused the stock price to rise rapidly. A Tweet, about 30 seconds later, at 11.41 AM, by 'zerosum24' confirmed that the rumor had reached Twitter and people had started talking about it. The sentiment had started rising rapidly by this time. The changes in S-Score™ and S-Delta™ were significantly positive. At 11.42 AM, the sentiment was over 2, and was statistically significant.

It was soon realized that this might be a hoax and that no offer was made. At 11.42 AM, TurboResearch questioned the credibility of the buyout offer.

There had been no official statement from Bloomberg, and hence, both the sentiment and the stock price kept rising. At around 11:50, a journalist from Bloomberg Tweeted that the news was a hoax and that it was not to be believed. This is the point where the sentiment started declining when people starting tweeting negatively and pointed out that the rise in stock price was a result of a rumor. The stock price dropped rapidly, but the sentiment didn't drop to its lowest point, since people who hadn't come across Bloomberg's and Twitter's denial of buyout claims were still tweeting about the buyout.

After that, there were mostly negative comments owing to the fact that the news was just a rumor. This led to the sentiment's being negative at the close. Figures 6a and 6b below shows that the SMA sentiment factors predicted the stock price quite accurately.

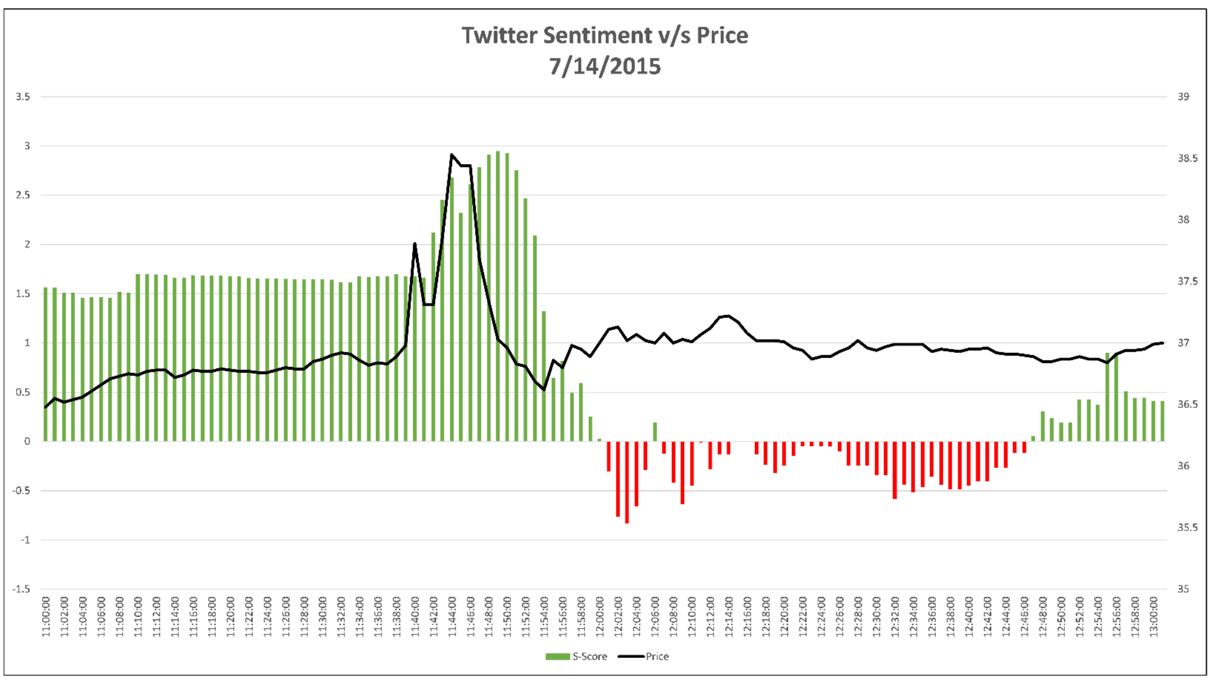


Figure 6a: **TWTR S-Score™ vs. Price**

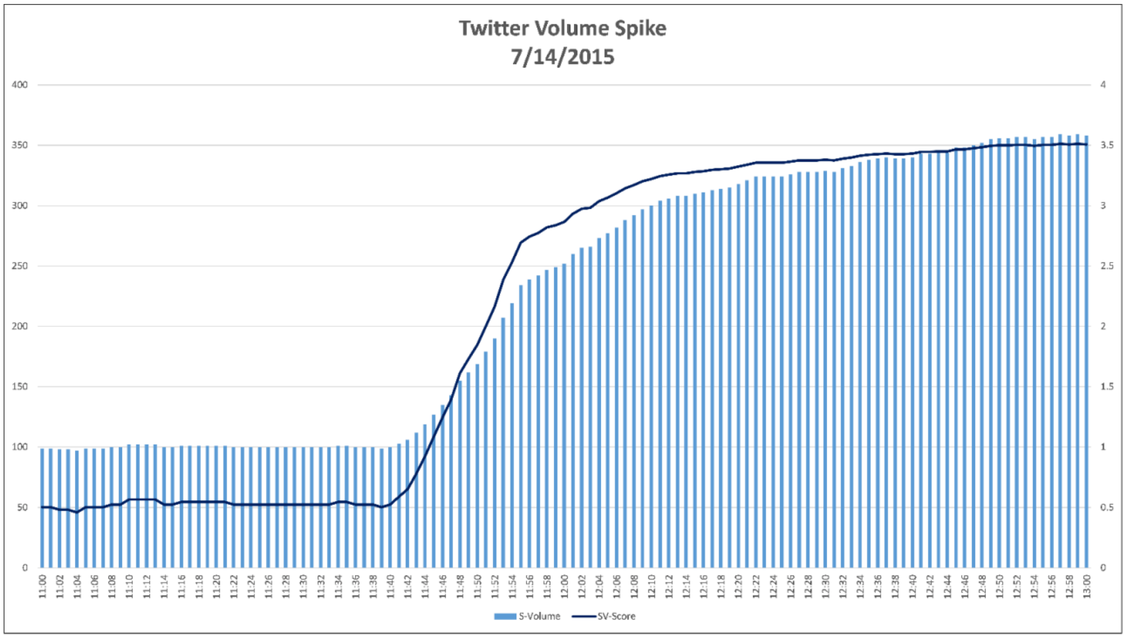


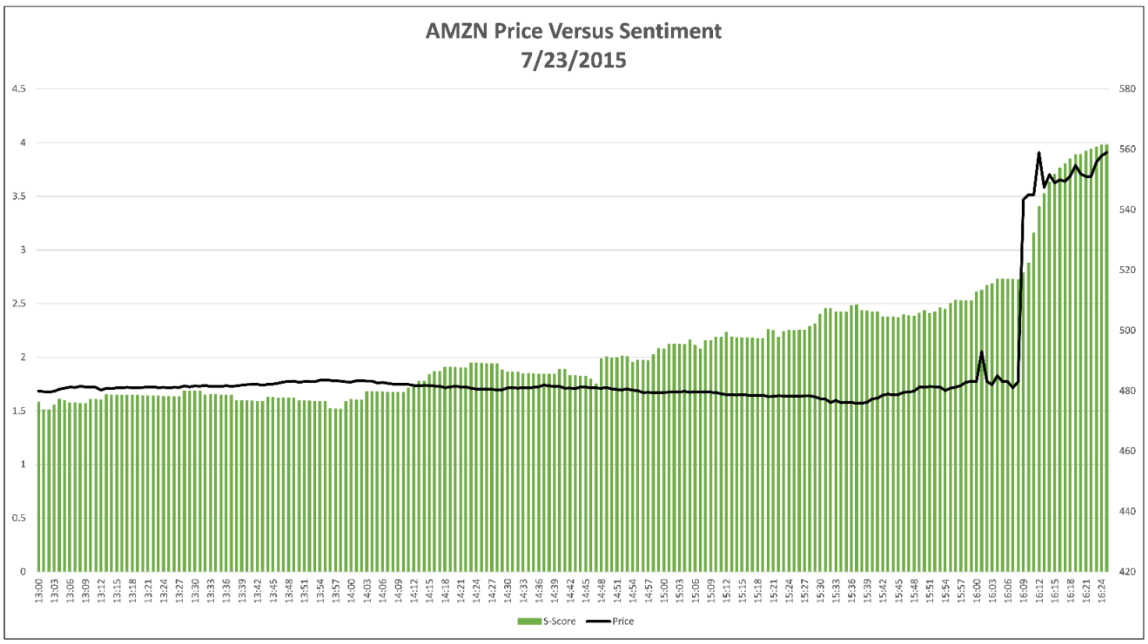
Figure 6b: Intraday S-Volume™ Chart for TWTR

High Profile Earnings Announcement: Amazon (AMZN)

Twitter sentiment can predict stock changes even after market close, as in the case of Amazon. Amazon reported earnings on July 23, 2015. While the market consensus was that the company would not beat expectations, the conversation on social media was different.

SMA data showed a sharp increase in sentiment metrics around 2:49 PM EDT. By 2:51 PM, the sentiment on Amazon was two standard deviations higher than its typical level. The stock was trading at $480.45 at this point. At market close, it traded at $482.18, higher than the price at the time when sentiment on Amazon became positive.

It was interesting to see how the stock traded after-hours once the company reported earnings. Amazon's stock shot up more than 17% -- to $568 -- from its price at 2:51 PM EDT after the company reported a surprise quarterly profit. The hidden sentiment value in Twitter data predicted what "conventional" market speculators failed to predict.



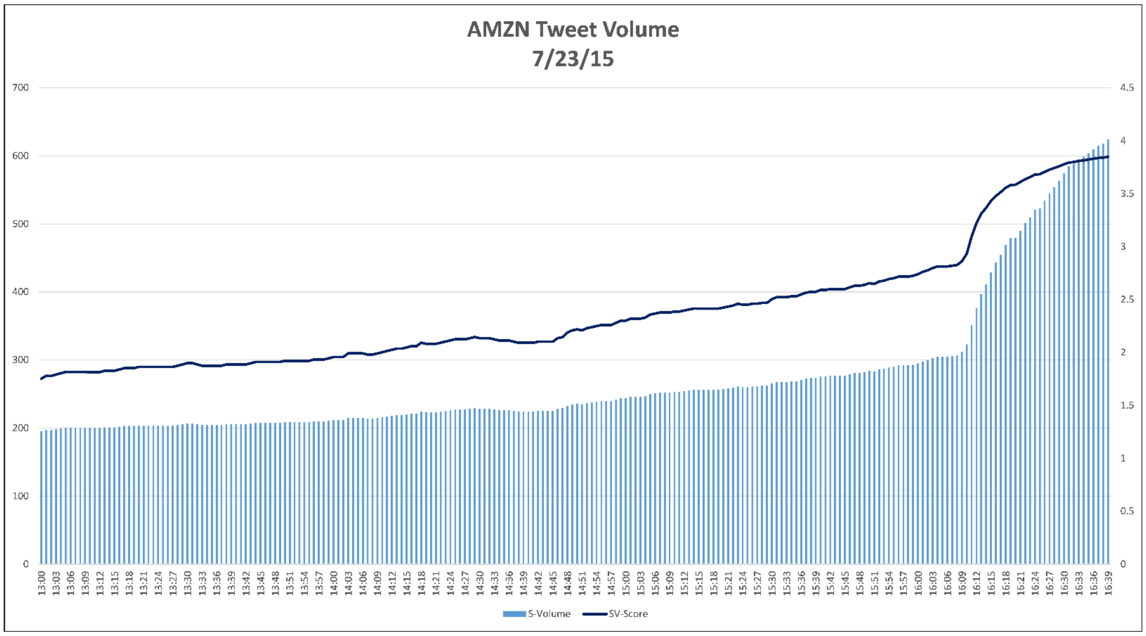


Figure 7: Intraday S-Score™ And S-Volume™ Behavior across Amazon's Earnings Event.

The progression of intraday S-Score™ and S-Volume™ metrics for Amazon is shown in Figure 7 from 1:00 PM EDT to 4:25 PM EDT. Amazon's sentiment remained positive throughout the day and became significant around 2:50 PM. The sentiment saw a sharp rise post the earnings announcement after market close.

Heat Maps for Stocks in SMA Universe based on S-Factors™

FIG. 8 demonstrates the use of SV-Score, the z-score normalized representation of micro-blogging volume, to reveal features of signatures embedded in heat map visualizations of market trading sentiment. Chart 1 shows the current practice to construct a market sentiment map, observable at sites such as StockTwits.com. In this visualization, a stock's color intensity level is mapped to some measure of trading sentiment for the stock, or to the stock's change in market price, or to the stock's share volume. The area on the map allocated to a particular stock is proportional to the raw number of micro-blogging messages observed for the stock over an observation period. Thus, the stocks with the largest total number of messages will occupy the largest areas on the map. Chart 1 shows a typical scenario, observed on May 25, 2012, in which the stocks such as AAPL, GOOG, JPM, AMZN, and MSFT occupy large map areas, but have moderately positive, moderately negative, or neutral sentiment levels. Stocks such as AAPL, GOOG, and AMZN will always dominate the current visualization practice because these stocks consistently have high raw message volumes, each day, irrespective of sentiment content. Thus, the current practice has limited usefulness to detect significant changes in sentiment signatures. On that day, stocks such as KLAC and ULTR had extreme sentiment levels and significant changes from normal levels of message activity as measured by SV-Score, but are obscured simply because these had far fewer total number of micro-blogging messages compared to the dominate stocks.

The market map in Chart 2 shows the result of re-scaling where the areas are a function of a stock's SV-Score metric. In this representation, a number of stocks with extreme sentiment levels and unusual social media activity, such as KLAC and ULTR, are detected and emerge from the clutter of Chart 2. JCP was retained in the re-scaling Chart 2 because the stock had extreme negative sentiment <u>and</u> unusually message volume on May 25. The high message volume stocks of that day, such as AAPL and JPM, occupy regions appropriate for their sentiment levels and historical normal message volume levels.

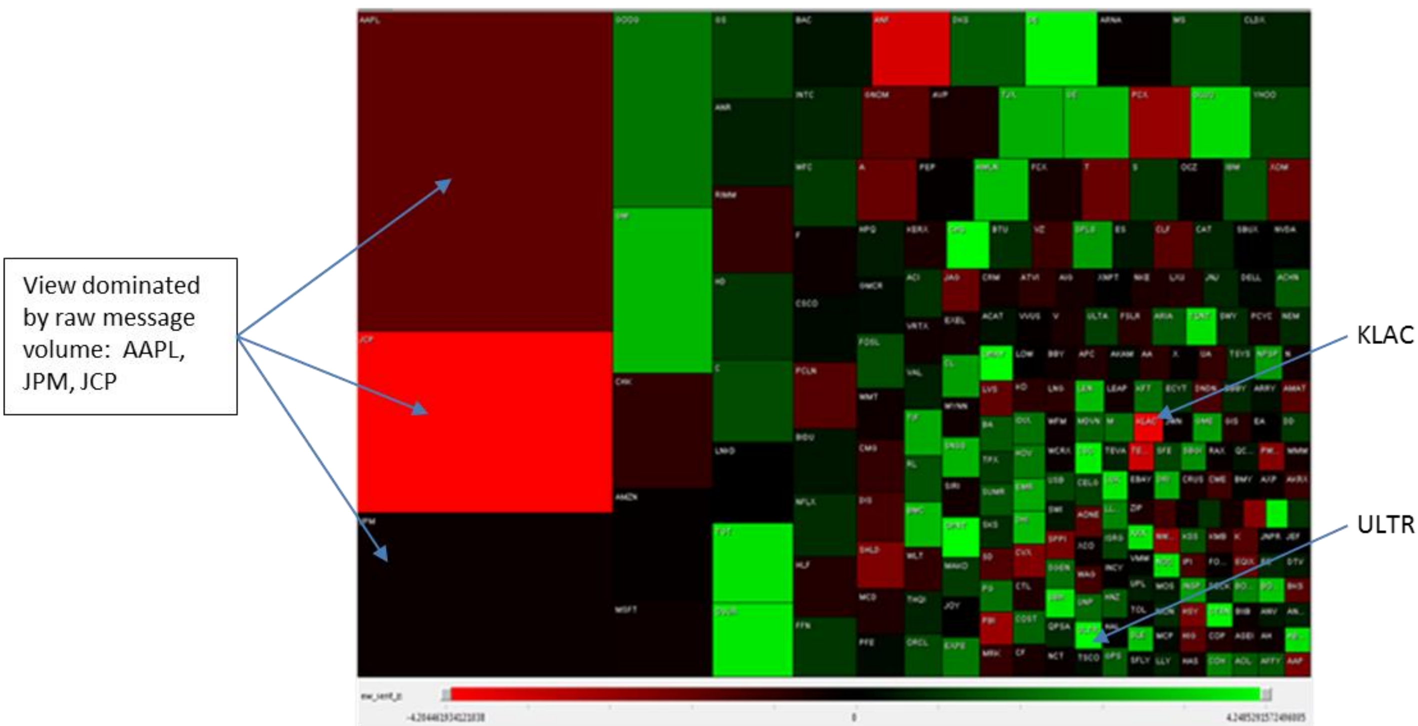


Chart 1: Heat Maps based on S-Volume and S-Scores

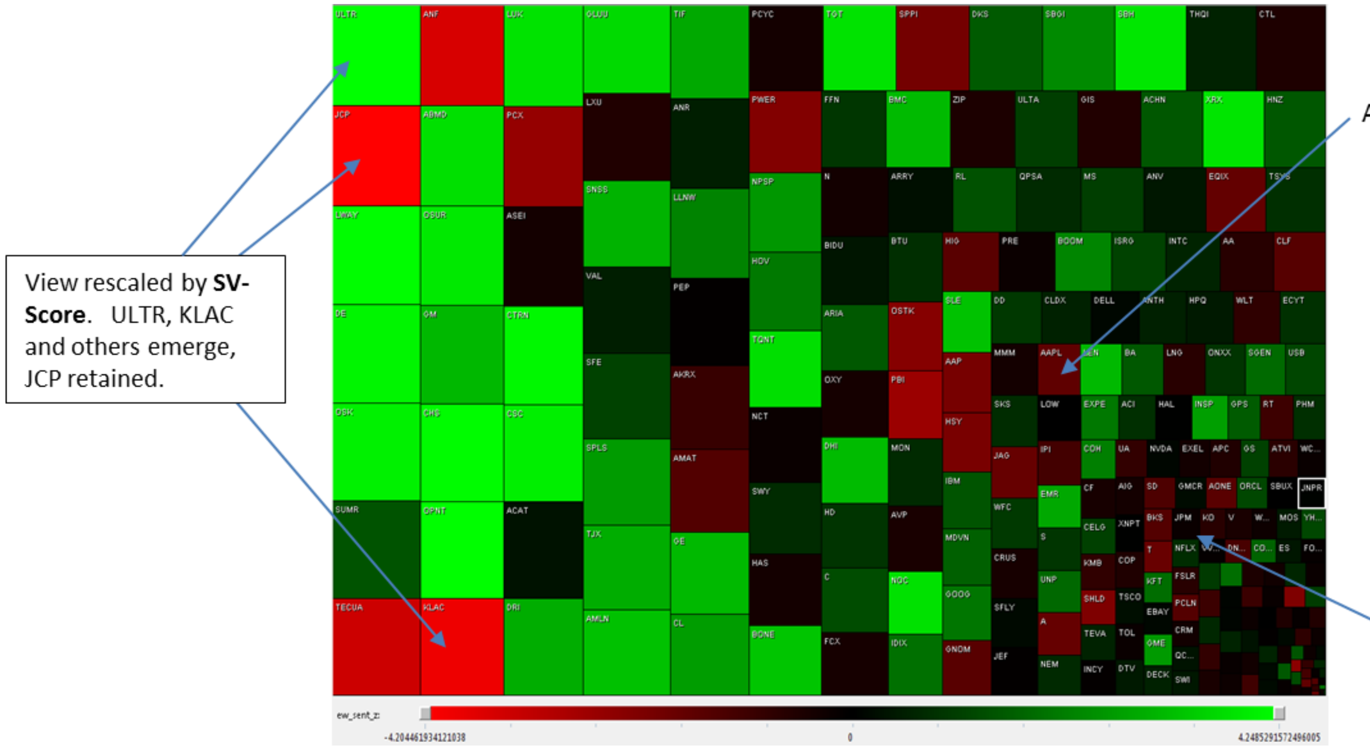


Chart 2: Visualizations Rendered With Volume Compared To Change in Volume

Figure 8: Market Heat Maps (5/25/2012).

Concepts for Applications

Upon request we make historical samples of our data available for download. Users have the opportunity to assess for themselves the value of the data we provide. Our documented research and analysis, and other independent work, have attested to the innovative added value S-Factors™ offer. Below are some points and ideas to consider in your analysis.

1. S-Factors™ exhibit different behavior at different sentiment levels. This is not surprising and is a testimonial to the embedded value in the data. It is highlighted in our earlier scatter plots analysis.

2. Daily S-Factors™ are categorized in three regimes: pre-market, post-market, and intra-day. Each warrants its own assessment.

3. Market events such as earnings reports, mergers, and acquisition announcements present good case studies, but tend to overshadow social media sentiment. It is interesting to assess S-Factor™ relevance prior to, and/or, after such market events.

4. S-Factors™ have predictive power for other market data, in particular options volatility.

5. S-Factors™ are published 24/7. Although weekends in general exhibit slower activities, it is possible to detect activities over the weekend which tend to be very informative.

6. S-Factors™ are useful elements in regression models to predict volatility, asset returns, and volume on daily and intraday timescales.

We publish our own research and analysis. We invite you to check our Research site for new updates and publications

About SMA

Social Market Analytics, Inc., (SMA) is focused on converting large-scale data into actionable business intelligence. Our data professionals have over one hundred years of combined data modeling and analysis experience. We have extensive experience programming databases and creating analytics for financial market clients, including hedge funds, money managers, and investment banks.

SMA offers a full range of large-scale data analysis services. We have experience with all major database technologies and analysis environments. We combine our extensive analytics capabilities with the latest technologies to assist our clients in the optimal use of Social Media data, tailored to their business needs.

Contact

For questions regarding the S-Factors™ concept and associated data, please contact:

Aditya Sharma: Adi@socialmarketanalytics.com

Jeff Blaschak: jeffb@socialmarketanalytics.com

Support: Suport@socialmarketanalytics.com