News Based Stock Sentiment

News Based Stock Sentiment
Viv Penninti and Ravi Koka, Stocksnips, Inc.

In November 2016 we had published a research note "Stocksnips Sentiment Analysis Research" where we examined whether News Sentiment was a leading indicator of stock price changes. That study was limited to a single stock AAPL and the results were promising. We have now extended our research to cover a larger number of stocks and two years of historical data. In this paper we report our findings which demonstrates the construct validity and reliability of news sentiment. Overall, the Stocksnips sentiment signal (V3) provides statistically significant ($p < 0.01$) increase in stock price correlations – using better message weighting and attenuation heuristics

1.0 Background
Alternative data sets – in particular stock sentiment signals – have been widely mentioned in the financial industry over the past few years. While sentiment is expressed and understood normally as a % positive indicator (whether at the stock ticker or ETF or Sector level), the manner in which these signals are derived is vastly different. Examples of the data used to derive sentiment include the following:

- One class of sentiment signals uses news articles, blogs, and other textual information to classify an entire "object" (e.g. the entire article or blog) using the Stanford classifier or other sentiment classifiers (which are not trained per-se with financial nuances/terminology in terms of assessing sentiment, but rather use "classical bag-of-words" to assess sentiment). These models tend to suffer from a lack of precision and noise due to (a) the unique nature of financial data and terminology, and (b) due to the aggregate nature of the assessment at an article level – which by itself is a

collection of statements some of which are positive and some could be negative. The Stocksnips approach is more sophisticated and uses the same input but the attribution and classification engine, as further explained, is much more discrete and granular and trained to handle financial news (unlike using a generic Stanford classifier).

- Another class of sentiment signal is derived almost exclusively from social media – aka "social sentiment" which are touted as "crowd sourced" sentiment indicators. Vendors with these signals rely on "volume" and "velocity" of social comments as the value creators in terms of reliability and construct validity. Research, however, has shown, that aside from the truly large tickers (AAPL, FB, AMZN, etc.) tweet volumes for 90% of the tickers tend to be relatively small – and highly noisy and volatile. Furthermore, social signals beg the question: are 1,000 tweets worth a single article from an experienced analyst? What is the memory of user-based tweets (i.e. is there message value that decays over time– or are they truly independent readings for each day)?

- A third class of pseudo sentiment signals is derived from financial statistics such as a "10-day equity only buy to open put/call volume ratio". This measure is the number of puts bought to open, relative to calls, over a roughly ten-day trading period (from the International Securities Exchange – ISE). These signals use purely financial indicators – and have nothing to do with analyst or news articles. While they serve a purpose and have been shown to have value, they should not be compared to sentiment derived from news articles and other publicly available textual information (e.g. SEC filings, Earnings transcripts, etc.).

To summarize there are numerous "sentiment" signals being used in the financial industry but each of them measure something different – i.e. they are not really comparable. Furthermore, the reliability and construct validity of such signals is unclear and opaque due to the nature of the data collected and the inherent challenges to ensuring a meaningful sentiment signal (regardless of velocity and volume factors).

2.0 Stocksnips News Sentiment Signal
Stocksnips has developed an AI based News Sentiment estimation system which uses news articles from established sources and asserts the sentiment of each financially "oriented" sentence within an article. Using AI driven attribution and Machine Learning models trained with large labeled data sets specifically developed for financial sentiment – the Stocksnips AI engine is unique in terms of both its approach and rigor. The aggregate value of news-based sentiment assessed at the discrete sentence level using the financially trained AI model, from reputed news sources, can a-priori be expected to be less "noisy" and possibly less "biased" versus social and other types of sentiment signals.

Assessing sentiment at the snippet level is only the first step in the process of generating a sentiment signal. The industry standard practice would be to compute a simple ratio of Positive Count divided by Total Count of snippets for a ticker (or possibly the positive score/total score). We note, however, that "Count" of snippets is a "discrete" variable. Furthermore, we know small stocks do not have as many articles resulting in lower number of snippets. The distribution of such counts is "non-normal" due to this low frequency. The distribution of snippets is more like a Poisson distribution with lambda = average snippets/day < 1. However, the Poisson requirement for independence of events (if these events occur with a known constant rate and independent of the time since the last event) – is unlikely to be true – which is generally true of many real-life situations. This type of occurrence where events occur in clusters could also imply "auto-correlation" – in which case the Poisson process could be modeled using Markov chains. Other factors to consider when creating a sentiment signal include the following:

- News Volume: The volume of news snippets on a given day can influence 'how much weight" to put on this informational content contained in these snippets. Below a certain threshold (say 10 snippets) – the current period news could have lower value due to sample size issues for example. Even in the case of high volume, the importance of "recent message" history will need to be considered.

- News Frequency: The frequency of news for a given ticker – i.e. the number of days between snippets is also another factor to consider. Average news volume for example could be high (based on a SEC 10k filing for example) – but if this "news" is only available once a quarter – how is that different from a stock with the same news volume average but with higher frequency?

- News Sources: The source of news articles is not "homogenous" – the "raters" (the analysts writing the articles" are not the same) and the articles themselves are not continuous sources of information (e.g. SEC filing versus non-SEC news). The discrete and varying nature of the sources confounds and complicates the creation of a reliable measure which has construct validity.

- Decay of News: This factor relates to how news (snippets) decays over time. Clearly old news (say some t=360 days ago) in some cases should have zero impact on today's sentiment – while more recent news has the highest impact. This "attenuation" of message weight over time is a critical factor for ensuring signal reliability.

- Market Factors: This includes cases where there is a systemic change to market sentiment that affects all stocks. For example, macro-economic factors such as an oil embargo or interest rate changes could affect the entire market – and may not be reflected in individual stock sentiment at the same time. The nuance here is that when there is minimal snippet (news)

information for a stock, then we would need to correct for this stock's sentiment using an external market factor.

Using traditional indicators such as 7 Day moving or 30 day moving with such discrete data (at the ticker level in particular) is clearly problematic especially where "lambda" or the mean snippets per day is small or has low frequency and from a statistical perspective "reliability" and "construct validity" need to be reviewed and addressed.

3.0 Assessing Signal Reliability
To review the performance of the new sentiment signal, we assessed two key aspects: (a) reliability, and (b) construct validity – as further explained below.

- Reliability: "A measure is said to have a high reliability if it produces similar results under consistent conditions." In the case of stock sentiment, a question to ask is whether the variability of stock sentiment "makes sense". Is there "random variation" in daily sentiment that is unlikely? Clearly if sentiment is smoothed (or averaged over some period of time) – it may appear to be less variable but then does it reflect current sentiment? Note by definition a moving average of duration "N" – includes data from N-1 periods ago and there is historical "baggage". What is the "attenuation" level of historical values? Fundamentally the reliability of a signal cannot be confused with reliability of a smoothed signal – since smoothing could in effect reduce construct validity.

- Construct Validity: This can be defined as "the degree to which a signal measures what it claims, or purports, to be measuring". If we are measuring true stock sentiment then it should be correlated to stock price and possibly weakly correlated to market sentiment among other factors – which could include correlation to trading volume

4.0 Sentiment Signal Study Results
Stocksnips conducted a study to assess the reliability of its sentiment signals using a randomly selected sample of 629 tickers from its database of 3,000+ tickers with sentiment information. The study period covered from August 2016 to September 2018.

4.1 Definitions
A few key definitions used in the analysis include the following measures:

- 7D % Positive Sentiment: This is a sentiment signal (original version 1 release of Stocksnips) wherein sentiment was derived as the ratio of seven days total of positive snippets divided by the seven-day total of all snippets.

- Daily Sentiment V2: This is the sentiment signal (version 2 release of stocknsips) which uses an exponential smoothing model on positive and total counts to derive the sentiment ratio/signal.

- Daily Sentiment V3: This is the latest sentiment signal (version 3 release of Stocksnips) which uses optimized weighting for message decay factor, message volume, and source of news weighting).

The issue of correlating stock price with sentiment over a long period of time is inherently complicated since sentiment is a percentage metric while stock price can increase or decrease "without limits" – a sentiment score is "bound" between 0 and 100%. Hence a simple correlation of stock price versus sentiment score **over a long period of time** where there is significant stock price movement would not yield a meaningful correlation. It is for this reason that correlations to assess construct validity were conducted using deviation from a baseline (50 day moving average).

- % Price Change: This is the ratio of current closing price divided by the 50-day moving average price minus 1. This primarily provides the percent change in current price over a 50-day moving average price.
- Sentiment Deviation: This is the difference between the current day sentiment minus the 50-day moving average sentiment.

The above measures were used for all correlation calculations in this document.

4.2 Descriptive Statistics

As mentioned earlier, a total of 629 tickers were selected for this portion of the analysis. The average ticker had 1,110 snippets over the study period which translated to an average of 1.79 snippets per day. The average ticker had news only 17% of the days (including week-ends) – and the average market cap of the sample was 16.7 billion.

| Measure	Total
Count of Tickers	629
Average Snippets per Ticker	1,110
Average Snippets/Day/Ticker	1.79
Average % Days w/ News/Ticker	17.0%
Average Market Cap ($ Billions)	16.7

4.3 Ticker Sentiment Reliability Analysis

The intraclass correlation coefficient (ICC), is a statistic that is often used to measure "reliability" of a metric. Conceptually, reliability for a quality measure means measuring the same data twice and detecting the "reliability" of such scores.

| Sentiment Measure	ICC	Standard Deviation
7D % Positive Sentiment (V1)	0.34	15.6
Daily Sentiment - Smoothing Model (V2)	0.47	11.2
Daily Sentiment - Enhanced Model (V3)	0.63	5.0

The ICC scores ranged from a low of 0.34 (for 7 Day % Positive measure of sentiment) to 0.63 for the V3 model with message weighting. One can also observe a significant reduction in daily sentiment variability with V3 (which if construct validity is unchanged would suggest a more reliable measurement) and much higher ICC – indicating higher reliability of this measure versus the other two sentiment measures. Since the sentiment scores are being compared across time periods (within effects) the assumption of similar measurement across time periods need not be true – and hence the measured ICC may not be an accurate indicator of reliability.

4.4 Ticker Correlation Analysis

The correlation of % price change of a ticker (close price/50 day moving average price – 1) versus sentiment deviation (current day sentiment – 50 day moving average sentiment) was computed for each ticker to test for construct validity. The market cap weighted correlation across the 629 tickers was found to 0.23 (for version 3 model) versus 0.19 (for version 2 model) – which is statistically significant at $p < 0.05$ level. This 20% increase in correlation along with the reduced variability (or higher reliability) indicates that the new sentiment model is more robust and useful than the older model.
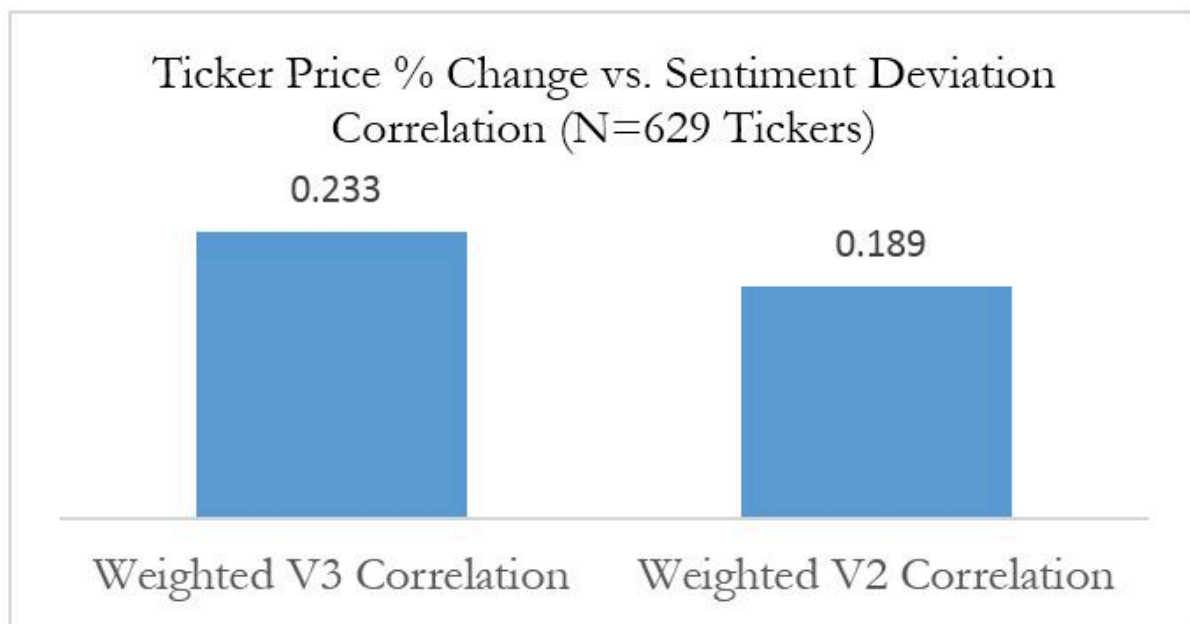
There is not a significant difference in the number of tickers which show a positive correlation with V3 at almost 74% while V2 shows 72%. If the null hypothesis is that the sentiment signal is white noise, we would expect 50% of the tickers to show positive correlation.

% Companies with Positive Correlation (N=629 Tickers)

73.7%   72.2%

% Companies w/Positive Correlation V3   % Companies w/Positive Correlation V2
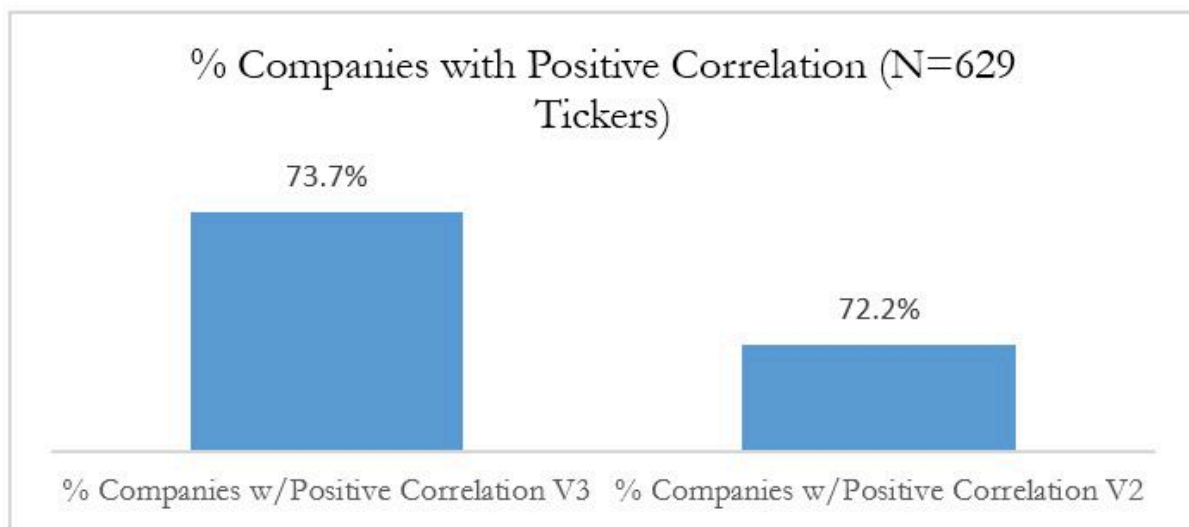
There appears to be a clear difference in terms of some tickers showing negative correlation having a lower market cap and a lower number of snippets per day (lower amount of news). The market cap weighted correlation for the 74% of positive only tickers is approximately 0.30 which is much higher than the overall average of 0.23.

Companies With	Count of Tickers	Average Snips/Day	Average Market Cap (Billions)	M-cap Weighted Correlation
Negative Correlation	164	858	$ 9.8	(0.13)
Positive Correlation	465	1,198	$ 19.1	0.30
Total	629	1,110	$ 16.7	0.24

Further analysis of market-cap impact shows that there is a significant decrease in correlation between price and sentiment for very small market cap companies (less than 0.30 Billion). This is most likely due to negligible news other than quarterly SEC filings (which is unlikely to have any impact on price beyond the announcement window).

Market Cap	Correlation Index	Tickers
<0.30B	43%	100
0.3B-3B	81%	241
>3B	100%	288

## 5.0 Sectoral Results Using V3 Sentiment Signal

With the observed increase in reliability and construct validity for the new version 3 message weighting model for sentiment, the model was applied to all 3,000+ tickers in the for all historical periods beginning in January 1, 2016. This database was then mined to review correlations at aggregate – sector levels. Sector level closing price was derived by weighting each ticker in a sector with its market cap and the same was done with sentiment. A few sector level charts are shown below:

All Sectors (All Tickers)
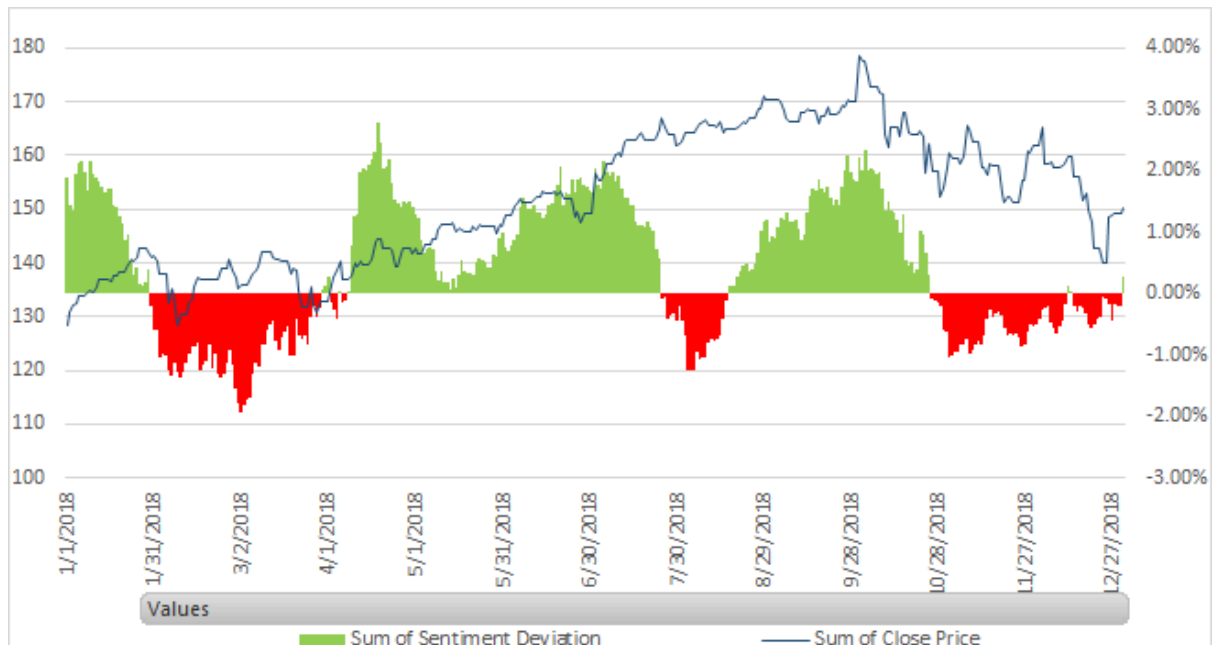
Above chart shows a clear correlation between market weighted closing price on the left axis and sentiment deviation (from 50day moving average) on the right axis.

Technology Sector
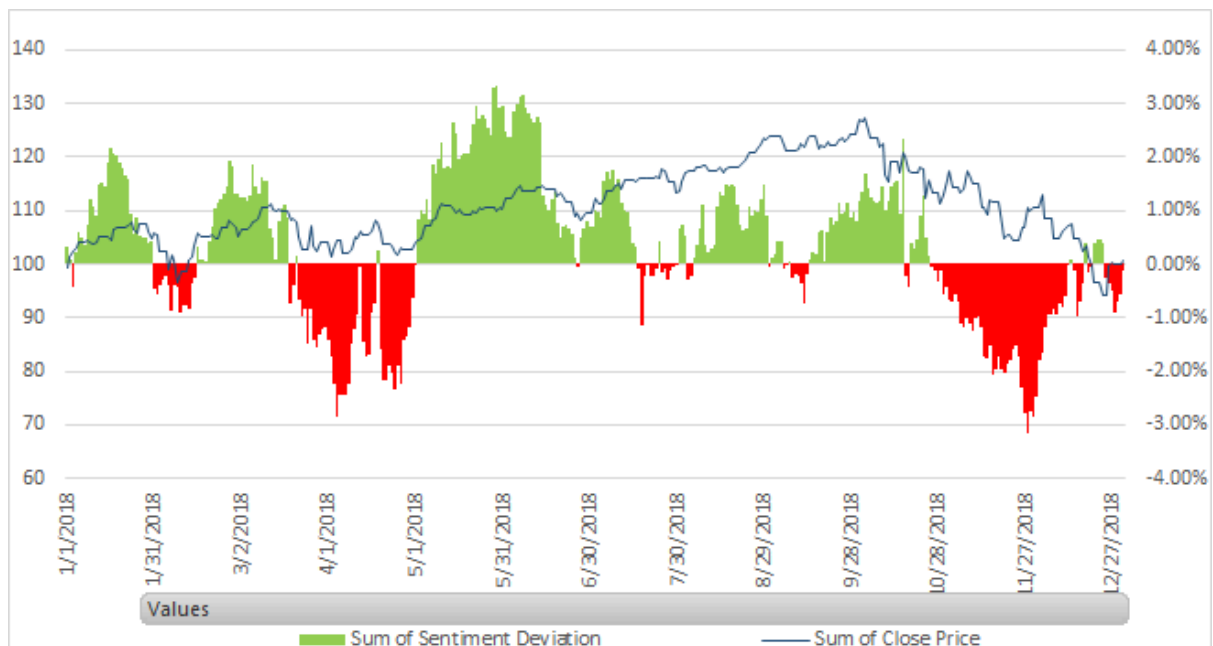
The technology sector chart also shows a strong correlation as seen above.

A summary of the estimated correlation (using 2018 trading days only) – for each of the sectors is shown in the table below.

	% Price Change Correlation vs.		
	10 Day Sentiment Deviation Lag	5 Day Sentiment Deviation Lag	Current Sentiment Deviation
All Sectors	0.36	0.39	0.42
Communications	0.16	0.21	0.38
Consumer Discretionary	0.22	0.13	0.08
Consumer Staples	0.19	0.30	0.37
Energy	0.35	0.38	0.45
Financials	0.07	0.06	0.14
Health Care	0.21	0.26	0.29
Industrials	0.29	0.28	0.29
Materials	0.30	0.37	0.39
Technology	0.42	0.59	0.65
Utilities	(0.16)	(0.12)	(0.08)

Overall market correlation with sentiment deviation is 0.42 – with the correlation dropping to 0.39[1] with a five-day sentiment deviation lagged and 0.36 for a 10-day lag. Technology sector appears to have the highest sectoral correlation at 0.65 (possibly related to both the volume and velocity of news for this sector). The low negative correlation of the Utilities sector is unclear by may be related to the counter cyclical nature of these stocks. Overall, the level of correlation for many of the sectors and total market is strong with a reasonable lead effect in some cases.

6.0 Conclusions

It is known that stock price is impacted by many endogenous factors such as earnings per share, market share, growth, etc. and exogenous factors including macro-economic factors, political events and "random white noise" events. Historically, and even today, prediction of stock price movement is inherently complex. However, in a globally connected world there may be an opportunity for information arbitrage. Companies are tied to other companies and large conglomerates (such as Apple) can impact the performance of other companies. If there is asymmetry in timing of information published by these interconnected companies, it is posited that such (news) information can influence stock movements. The results using the Stocksnips latest sentiment signal with optimized message weighting show (a) higher reliability of this signal (versus most other signals including social sentiment signals which are notoriously volatile), and (b) statistically significant correlation between sentiment changes or deviation and stock price changes. This effect is even more apparent at the sectoral level as is to be expected. Including Stocksnips sentiment signal in portfolio allocation models and buy/sell models as an alternative signal is highly likely to increase "alpha" due to the findings mentioned in this report.

[1] Note: Lag effects are much more pronounced at the ticker level. 5 day lagged correlation at the ticker level appears to be optimum from a correlation perspective.

Privacy Policy