Getting Ahead of The Crowd !! - A Visualization of Price Response to Social Media

(Aditya Rajmane (anr331) , Mark Weisenborn (mw1556) , Monil Suthar (ms8624))

In this update, first, we address your question and those of the instructor.

… I would like to see this addressed in the next update you submit, is the <u>*related work section*</u>*. What we were expecting in the related works sections is a discussion on what similar/relevant work exists in the space. It may include research papers, articles, stories etc. Please remember that the related works section usually help you discover what may or may not be the real contribution of your work, and whether is it a novel work or an implementation of something that already exists. I will be deducting 2 points for not including relevant related work.*

Here is relevant related work:

We located a research paper that explored how the frequency of positive and negative words influenced the prices of mentioned securities. This is available here and is fully described within the material below: http://www.bhwang.com/a_research/z_papers/5_wisdom%20of%20crowds.pdf

Figure 3. Seeking Alpha and Abnormal Returns over Different Holding Periods

This figure reports coefficient estimates from regressions of abnormal returns on measures of the views reflected in *Seeking Alpha* (SA) articles and comments. The sample period is 2005-2012. Abnormal returns are the company's raw returns minus the return of a value-weighted portfolio with similar size/book-to-market/past return-characteristics. The horizons over which cumulative abnormal returns are computed are 1 month, 3 months, 6 months, 12 months and 36 months. The regression equation is identical to the one in column (3) of Table 4. Here, we plot the coefficient estimates on $NegSA_{i,t}$ and $NegSA\text{-}Comment_{i,t}$ along with their corresponding 95% confidence intervals. Standard errors are clustered by firm and year-month.

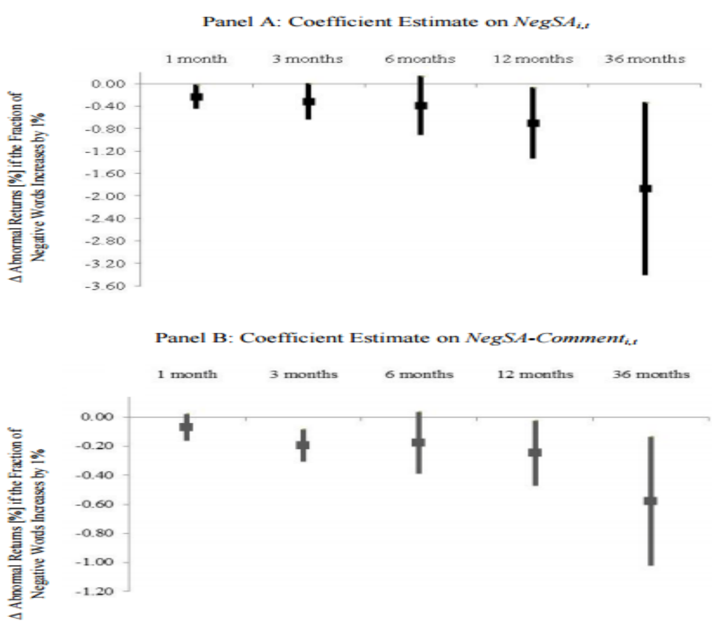
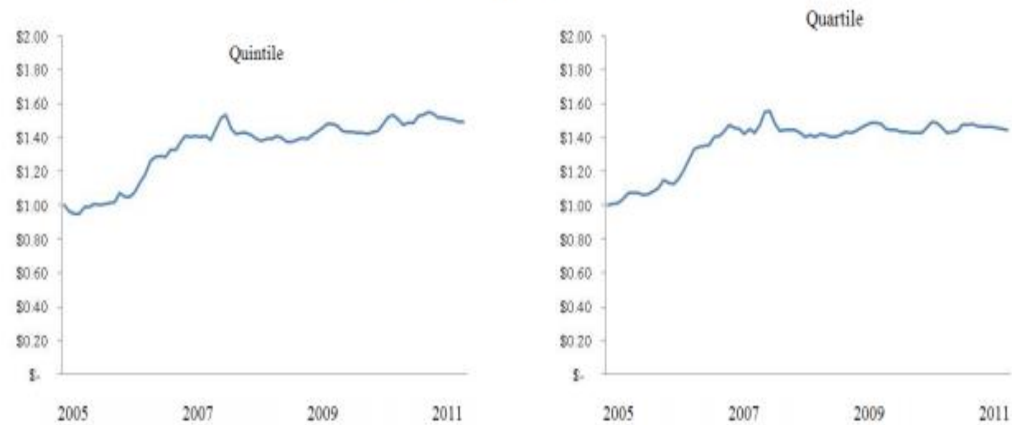


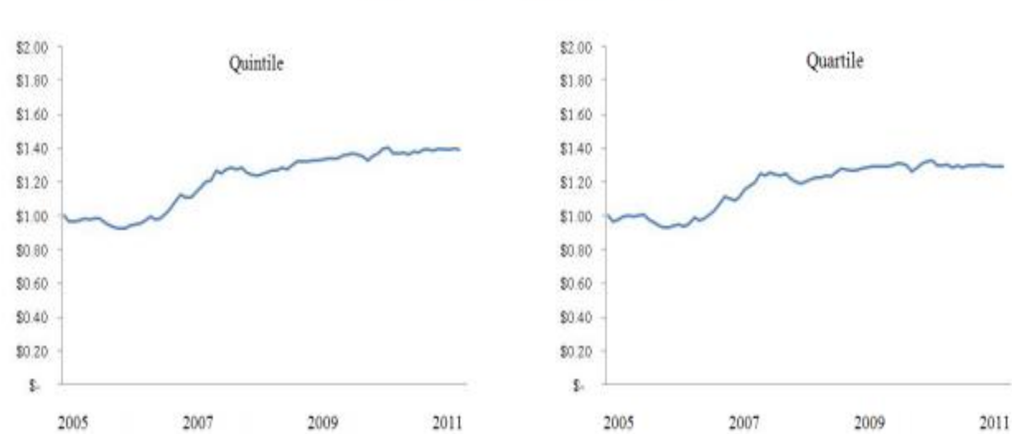
Figure 4. Seeking Alpha and Abnormal Returns over Different Holding Periods

This figure depicts how \$1 invested in a simple calendar-time trading strategy would have evolved. The trading strategy is as follows: At the end of each trading day t, we assign stocks into quintile (quartile) portfolios based on the average fraction of negative words across all articles published on SA about company i on day t ($NegSA_{i,t}$); we also form quintile (quartile) portfolios based on the average fraction of negative words across SA comments posted over days t to t+1 in response to the SA articles ($NegSA\text{-}Comment_{i,t}$). We skip two days and hold each stock in its respective portfolio for three months. Based on the daily returns of a long-short portfolio, where we go long stocks in the bottom quintile (quartile) and short stocks in the top quintile (quartile)), we plot how much \$1 would have grown/shrunk through calendar time.

Panel A: $NegSA_{i,t}$ - Based



Panel B: $NegSA\text{-}Comment_{i,t}$ - Based



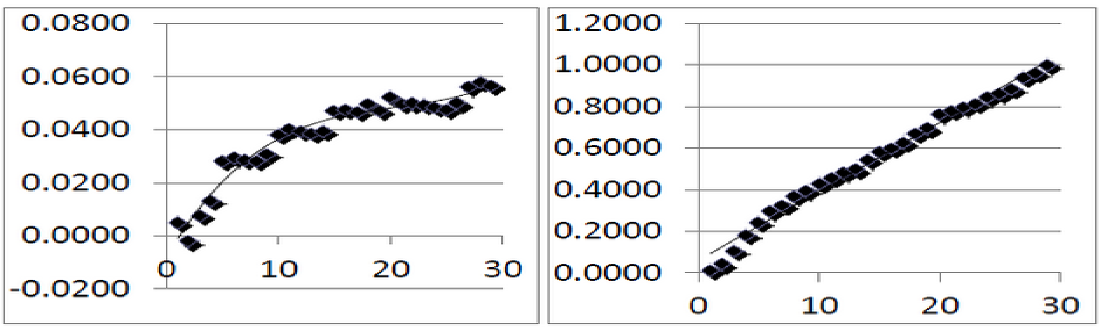
We also located additional work that shows intraday price movement following publication at a single news source, Seeking Alpha.   This study shows the price movement during the day, as fully described in the pasted body of this sample visualization.

Panel 1. Average Price Path and Average Cumulative Average Volume

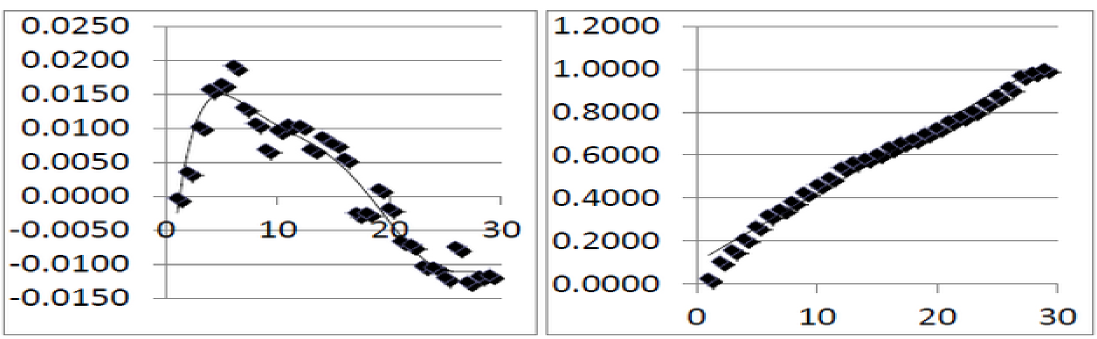
These figures present average trading price paths for the days of trading that were statistically significant. For each revision, price begins at time 9:30am and completes at 4:00pm on the x-axis and magnitude of change is presented in percentage format on the y-axis.

The left chart is price, the right chart is cumulative volume.

"Top" Micro Long Day t



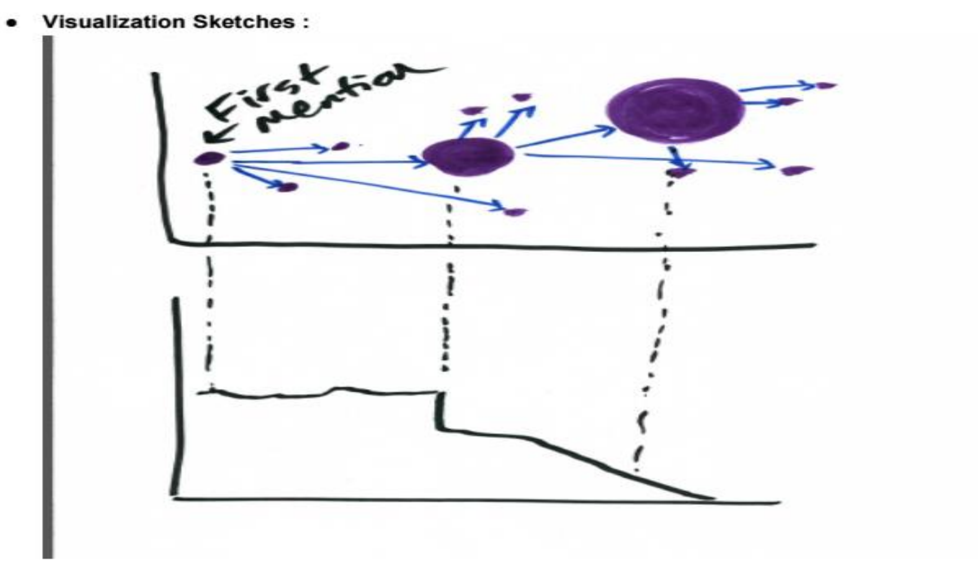
Micro Short Day t



Price and volume charts were constructed using tick data provided by Wharton Research Data Services. The trading day is constructed as a series of 15-minute bins beginning at 9:30am. Each point represents an average of prices or volumes closest to but not after each bin-time. For each category a small (<20) sample of securities that had tick data available from September to December 2013 was used to develop average values. These sorts of charts may be helpful for a trader trying to minimize implementation shortfall and to help understand the different price and volume behaviors of longs versus shorts. Because this data set is from the group that was announced outside of market hours the initial rise in price on the short chart may be due to closing a downward gap in the initial opening price print. In order to fully understand the price formation process every security should be included in the values used for depiction, a task beyond the scope of this academic study but worthy of future research.

Back to our work.

We feel we can do a better job than both of these pieces of relevant work by making a visualization that encompasses both **information spread and price response**, as below.   Look at how our **information is encoded** within the visualization – the arrows show flow of information (spread), the bubbles encode the size of the publication, and the dashes encode a common time between both visualizations. In the lower chart the single line encodes the price movement and we may add y-axis labels so the viewer can easily see the percentage change.



In this chart the information dissemination and price path is represented by arrows, balls, dashes, and a line. Larger balls are larger news sites and the dashes from the top chart to the lower chart depict the time when information was displayed (published) at each news site, and the corresponding price at that time. The arrows show flow of information (spread), the bubbles encode the size of the publication, and the dashes encode a common time between both visualizations.

We feel this visualization helps to answer questions related to whether size of news venue matters, whether a user can get ahead of price moves by paying attention to the smaller sources that feed to larger sources. We also currently working on mock-ups that incorporate “source traffic” and “story traffic” fields. We realize that the spread of news also relies on the fact that how much attention is a story getting on the web.

Task Abstraction:

(1)	Does publication at a small venue tend to be followed by larger venues picking up the story?

- The visualization shows the generalized links between news venues. We may add some single number above each arrow to show to frequency (as a percentage) of times a small news source preceded a larger news source publishing.

(2) Does publication effectively predict the direction and magnitude of the price change?

- The visualization shows time, ordering, size and resulting price move over a generalized day. We are using a random sample of the full data set per headline word to make generalized displays for how each word tends to propagate through the news sources, and the associated price response.

We also currently working on mock-ups that incorporate “source traffic” and “story traffic” fields. We realize that the spread of news also relies on the fact that how much attention is a story getting on the web.