Finance Bulletin

Finance Bulletin 1:1

Information, attention, sentiment, and buzz in the financial markets

Hubert de La Bruslerie[1]

[1]University Paris Dauphine.

Corresponding author: Hubert de La Bruslerie, hlb@dauphine.fr
Date of submission: 1/02/2017
Date of publication: 28/03/2017

Abstract

This paper aims at focusing on the avenues of research related to the process of information integration by taking explicitly into account investors' sentiment, investors' attention, and the buzz hypothesis. New social media introduce change in the way information is processes in the market. Qualitative concepts such as rumor, opinion, sentiment, are often put in the frontstage. Moreover the formal dimensions of information become more important compared to the content of information. This leads to new avenues of research aside the standard information value hypothesis.

1. Introduction

Efficient market hypothesis (EMH) asserts that financial market valuations incorporate all existing, new, and even hidden information, since investors act as rational agents who seek to maximize profits. Investors will follow an economic rationale and analyze new relevant information. They are able to separate information and pure noise. This process of integration qualifies valuable information in the market price.

However it is complex to analyze the information integration process resulting in an observable buy or sell decision at the investor's level. Traditionally information is supposed instantaneously incorporated. It is assumed that investors are continuously demanding information and able to process it. This infinite demand function is challenged by the limited ability to process information at the investor's level: Attention is a scare resource. Investors decide to allocate it or not to a list of specific stocks.

This process is conditioned by behavioral factors such as mood in the markets, herding behavior or overall attention in the market. At the end when the stock price moves, it may be due to new valuable pieces of information or to change in sentiment or to more/less attention in the market. These behavioral elements and these social factors may explain that the integration of valuable information may need time to be fully integrated in the stock price.

This paper aims at focusing on the avenues of research opened by the process of information integration in the market taking into account explicitly the notion of investors' sentiment, investors' attention, and the buzz hypothesis opposed to the information value hypothesis.

The introduction of new media such as Twitter and Facebook changes the way information is processed. These informational marketplaces develop interactions and rumors spread over easily and quickly. Following the analysis of McLuhan (1964), the volume of these new media becomes in itself the message. The idea of buzz illustrates this process of integration of information that develops in a circular and dynamic way between participants in an information marketplace where the producers are at the same time the users of information. We present the buzz hypothesis of information integration that covers at the same time the two opinion and attention mechanisms already identified in the literature.

2. Information processing and media

Financial markets rely heavily on internal institutions and actors whose function is to produce and disclose relevant information. Analysts are the best example. They produce highly synthetic information that is standardized and understandable to investors such as earnings forecasts and recommendations. The process of integration of these pieces of information is not so simple. Even if the content of the information is valuable it may be polluted by conflicts of interest. So the analysts system introduces by itself some additional uncertainty and/or some noise (Michaely and Womack, 1999; Dubois and Dumontier, 2006).

Strategic communication policies and hiring investor relation firms will also bias information delivery (Bushee and Miller, 2012). This will increase communication press coverage and disclosure activity; it develops dissemination to investors…At the end, it will reduce the bid-ask spread (Bushee *et al.*, 2010). But will it reduce information asymmetry? If disclosure is pure noise, the answer should be negative. If disclosure is private information, the answer should be positive. However , if the communication strategy repeats constantly the same (positive) new, it is more complex as the initial delivery of private information is no more private after that and becomes a quantitative communication strategy that aims at creating a buzz, i.e. at dragging overall attention to the firm's name.

Dissemination tools are very various: commercial wire services (such as Dow Jones Reuters, Bloomberg), analyst' meetings, issue/revision of forecasts, public media coverage, conference calls, and social media.

The specificity of social media

The Internet introduces new relationship with regard to information. Online generates massive datasets. Social media are Internet-based services where the content of information is freely created by the participants. Everybody can raise his voice, and give information, opinion and express postures. One important characteristic is immediacy of information. Avoiding the channel of traditional media and overpassing the intermediary role of journalists and editors, will result in communication and messages that are not filtered by ethical considerations, trustfulness and accuracy constraints. The two major players in the social media industry are Twitter and Facebook.

The content of information provided on large scale by social media may influence the setting of price in financial markets as well as traditional media and financial news services. McLuhan (1964) recalled that the medium is not a passive technology but creates a specific human environment and a "little world" context. It constitutes an active and conditioning process. A medium affects the society in which it plays a role, not only by the content delivered over the medium, but also by the characteristics of the medium itself. Quoting McLuhan, "the "content" of a medium is like the juicy piece of meat carried by the burglar to distract the watchdog of the mind."

What is definitely new with social media is the phenomenon of buzz or rumors. Buzz refers information created by the investor or the consumer. The potential investor/consumer itself is the medium. This medium is also the target of the communication; it is not an exogenous tool or a simple channel. The medium creates the content of the information by expressing postures and opinions. Medium and message are here interrelated. In that sense the McLuhan's proposition is verified. This is particularly strong when looking at social media. As a consequence the social media volume, or buzz, becomes more important than the message itself. What is more informative to the market is the loudness of the buzz more than the global sentiment or mood. This buzz hypothesis opposes to the opinion hypothesis as the latter focuses on the positive or negative content of the piece if information.

Social media like Twitter can play a role in financial market in different but complementary ways. It may raise awareness. It directs attention to specific stocks. This is also called the salience view (Sprenger *et al.*, 2014). It mays also inform retail investors by adding qualitative opinions or by associating postures to the information. This combines with a real time dissemination tool that multiplies the news coming for instance from third financial party.

	Underlying mechanism	Empirical measures
Attention	Draw investors' attention to certain stocks based on news (i.e. offer of additional pieces of information) or awareness (demand side)	Ex ante: search volume from Google Ex post: transaction volume
Sentiment	Mood or opinions about certain stocks	Number of occurrence of words belonging to a given list/dictionary, positive or negative tone
Buzz	Initiation of a dynamic of attention and circulation of individual opinions. Users and producers of information are within a social group.	Volume of messages initiated by and exchanged between consumers/investors. Facebook, Twitter.

Table 1: Concepts of attention, sentiment and buzz in the investors' information process

A lot of literature has focused on the role of Internet data in financial markets.

- A first strand relies on the content of news disclosed by Internet media and their subsequent consequences on stock price moves;
- Others focus on the idea of attention that is allocated by investors to financial markets and/or to a specific asset or stock;
- Numerous authors extract overall signals or sentiments from news media to predict future prices moves. They use search engine and language semantic analysis. The methodology and used to build sentiments should be adapted to take into account the immediacy of social media;
- In the wake of previous studies, the sentiments or opinions have been analyzed by associating them polarity and position (either positive or negative) with regard to a stock or an asset. Are these opinions lead information about the future returns of the stocks? Or is the reverse true?

The concepts of attention, opinion and buzz are close together and partly overlapping. Attention draws investors' resources of awareness to available information and further affects stock price by incorporating this information. But the sentiment channel seems may be unrelated to information or may be pegged with news, it is driven by individual opinions and/or investors' mood. The buzz channel plays a role similar to the attention channel. It generalizes it and spread it by combining opinions and mood among a social group.

3. Attention

Attention is a scare cognitive resource allocated by an investor that is aggregated through media which circulate information demand, and that is globally conveyed to a given stock.

The channels conveying information to the investors and then to the market after a buy or sell decision have been largely studied. Analysts are one of them. Other channels of information are important:

journalists, media coverage, and intermediary services providing news. When mainstream and well-recognized media diffuse corporate news, stock prices are particularly affected, even if the news were publicly known (Boulland *et al.*, 2016). This finding is important because it outlines the role of the medium. This way to influence investors can be strategically used by firms. Some diffusion technologies are more efficient in drawing investors' attention. This is particularly the case of English-speaking electronic wire services. The electronic technology will give more immediacy and the use of English language may insure a broader diffusion worldwide.

Boulland *et al* (2016) looked at European firms choosing to use international wire services. Increasing attention through the implementation of electronic English wire communication will yield a quicker investor response and so forth a stronger market reaction in terms of short term abnormal returns. Delayed reactions (i.e. stock price moves over the 60 days period after the new announcement) will signal inattention. Attention is measured by its result: abnormal increase in transaction volume. Boulland *et al.* (2016) show that firms that disseminate news through an English-language wire service obtain greater investor attention, compared with those that release their disclosures in non-electronic form and in a continental European language. Their study deals with a very standardized piece of information: earning announcement and its corollary earning surprise. The stocks receiving more attention are more actively traded and get larger immediate abnormal return. Electronic and English wire services are channels disseminating tradable information.

Investors' attention is related to a news announcement that provokes them. It stresses the effective demand context of the market. The market reaction is great in magnitude when attention is great. Conversely when inattention dominates the reaction is delayed (see DellaVigna and Pollet, 2009). Arditi *et al.* (2015) look at Google search and the magnitude of the stock price move by considering the absolute abnormal returns. They find a positive relationship.

Measuring attention can be done ex ante through the use of Google information queries rather than ex post through the effective transaction volume (Da *et al.*, 2011). When investors look for information using a search engine, information providers record frequencies and statistics. Google produces a Search Volume Index (SVI) which is available for given keywords. A strong relation between SVI changes measuring increase in attention and trading by retail investors is evidenced by Da *et al.* (2011) at market level. The attention theory of Barber and Odean (2008) explains an asymmetric effect from the retail investor's point of view. When looking at buying, the latter should allocate his scarce attention to a very large list of alternate investments. This list is quite infinite and exceeds his cognitive ability. When considering selling, the retail investor will only look at the limited number of assets he owns. Odean and Barber (2008) conclude that surge in investor's attention will lead on average to a buy and should predict higher stock prices in the short term (and possible correction in the long term). Using Russell 3000's stocks, Barber and Odean (2008) support this hypothesis with a rise of 30pb in return in the 2 week following the increase in attention. The attention theory (Da *et al.*, 2011) identifies a link between search volume in the Internet and positive trends in stock prices. More precisely a way to identify attention is increases in Google search queries.

Mao *et al.* (2011) used Google and Twitter to analyze both attention and sentiments. Looking at the former, they identify attention by the volume of queries containing occurrence of a dictionary of 26 financial words in the Google Insight for Search service (GIS). They cross the GIS query data with the volume of tweets mentioning the same word from the dictionary. The two search volume indicators coming from two different social media are highly correlated (+0.62). The Google volume of query is positively correlated with the stock volatility index VIX and the effective transaction volume.

A Granger causality analysis between the GIS data and the DJIA stock index is supported, with the Google search indicator leading the DJIA and not the reverse. The same result is supported when considering transaction volume: the GIS indicator leads the DJIA transaction volume for a 2 weeks period (and not the reverse).

4. Sentiments

"The sentiment theory predicts short-horizon returns will be reversed in the long run, whereas the information theory predicts they will persist indefinitely" (Tetlock, 2007).
Numerous researches have shown that volume of certain words in media associated with sentiment is a mood indicator for financial markets (Da *et al*., 2011). Mao *et al*. (2011) extract sentiment using the two words "bullish" and "bearish". Although this is quite a poor dictionary, they built a bullish-oriented Twitter investment sentiment, TIS. This TIS Granger-causes stock returns. This result supports the view that sentiments may lead markets prices moves. Tetlock (2007) has shown that high level of pessimism in Wall Street precedes lower market return in the following day. Tetlock *et al*. (2008) have also developed similar sentiment analyses at the firm's level. They demonstrate that a negative sentiment helps in forecasting lower firm returns. These two studies underline an important limitation linked to analyses of sentiment (and to analyses of attention) at the global market level. Looking at Dow Jones or Nasdaq indexes will aggregate many stocks and this may create an aggregation noise. For instance at the same moment opinions (or attention) with regard to a firm may be offset by an opposed move of sentiment (or attention) with regard to another. A firm's level analysis seems to be more relevant.

Methodology to build sentiment using social media

Different methodologies can be followed to build a metric of polarity between good and bad sentiments from tweets. They rely on Natural Language Processing (NLP, i.e. searching system to parse sentences in natural language such as English). It starts from linguistic rules or thesaurus to analyze corpus text using NLP that gives synthetic frequency information. To extract opinion from a text, two approaches are available:

- Lexicon based processing refers to the semantic orientation of words. These lexicons are exogenous to the problem, for instance the Harvard IV-4 dictionary used by Tetlock (2007) and Tetlock *et al.* (2008). These lexicons are publicly available.
- Classification learning methods will determine a classification process which is contingent to the addressed problem. The learning comes from a two-step process: a first step will calibrate the classification algorithms on word features that are known. A training dataset is used to classify sentences (here tweets). The scope and the quality of the training dataset are crucial as the predictive ability of the learning algorithms is only based on the training dataset. The methodology to calibrate classification algorithms can refer to either discriminant analysis, naïve Bayesian rules (Mao *et al.*, 2011; Divet, 2016) or neural networks (Heston and Sinha, 2016). These training sets are associated with manually analyzed sentiment opinions or polarity. The classification process of sentences should take into account that the vocabulary and syntax of tweets are specific. For instance the use of emoticons is frequent with ":)" introducing a positive polarity (and respectively ":(" , a negative one). The Go *et al.* (2009) training dataset used by Divet (2016) covers 1 600 000 tweets from which the later randomly extracts 200 000 tweets.

The learning process identifies reference "words" that are linked to either positive or negative polarity. For instance, Divet (2016) outlines that the word "condolences" has 30.8 time more chance to be correlated with negative sentiments than positive sentiments. Conversely "congratulations" or recommendation" are linked with positive opinions. The social media context imposes to scrutinize onomatopoeia and to integrate in the learned lexicon "nooooo", "booo", "ughhh" (negative polarity) or "hihi" (positive polarity). The lexicon classifier is contingent and may be influenced by the context or the time when it is elaborated. For instance the word "Verizon" is negatively correlated with sentiments in Divet (2016). It only means that the Verizon firm was prominent at that time. This is problematic since it implies that the brand name in itself is negatively loaded.

The second step consists in analyzing the tweets dataset based on extraction of tweets mentioning firm's name over a given period. For instance Divet (2016) considers 869 212 tweets referring to a firm belonging to the Nasdaq 100 index over the March-June 2016 period. A simple parsing method will first eliminate punctuation and sentences are considered as sequence of words. More sophisticated approaches will introduce syntax and grammar rules, and identify propositions and sentences. Normalizing will remove stop words such as "to", "a", or "the". Specific tweet semantic objects (emoticons) are included in the analysis. The "bag of words" approach considers tweets as unordered sequences of featured words or signs. This analysis is very rude and simplistic. It is however more relevant in tweets as the limitation of characters in each message leads to use association of word, signs and onomatopoeias and releases from the use of syntax rules. The remaining text corpus is then analyzed using frequency counting of featured words. As a result the accuracy rate of classification in Divet (2016) is 73.2% in reclassifying the training dataset.

The relation between sentiments on a firms belonging to the NASDAQ index and the Index return leads to aggregate the polarity of the tweets attached to each firm within hourly sub-periods. Aggregation here means to calculate the difference between the number of positive tweets and the number of negative tweets. This overall sentiment value is then normalized because there are a higher number of positive tweets compared to negative tweets (Divet, 2016; Heston and Sinha, 2016).

A methodological problem arises to follow homogeneous data point in the sentiment time series and markets data time series. Tweets are continuously exchanged. Extracting synthetic sentiment value implies to calculate at a given point the cumulated balance between positive and negative opinions. Previous studies have considered daily or even weekly sub-periods. More recent ones focus on hourly sub-periods covering the business day. This implies to consider similar time patterns when measuring the effective variation of individual stock prices or indexes during the business day. It introduces a cutoff hypothesis imitating tweets out of the open hour period (or allocating them arbitrarily to the price variation between the last day closing quote and the open quote). This issue is well known. Time series date assume that data are collected at regular intervals. This is not true as we have overnight and week end gaps. The question is sensible as the last transaction price takes into account the overnight gap and the potential arrival of futures tweets that are continuously produced.

Empirical results

To summarize sentiments, Sprenger *et al.* (2014) use tweets and naïve Bayesian classification to allocate them between positive or negative opinions. Their analysis is at the firm's level and is crossed with the firm's stock price return. Both Divet (2016) and Sprenger *et al*. (2014) have shown that sentiment will lead market price variations. Using a Granger-causality definition, tweet sentiments in the four hourly periods will help to forecast the Nasdaq 100 stock index hourly returns (Divet, 2016). However the lead signal is only significant at the 5%, which can be explained by the aggregation sentiment problem. Bullish tweets around an event will increase volume and yield positives CARs, and negative tweet spikes give abnormal high volume and negative CARs. Sprenger *et al*. (2014) show that volume is insufficient, and that sentiments contribute to explain stock market reactions.

5. Buzz

Mao *et al.* (2011) has shown that the volume of financial tweets words Granger-causes future returns overs a 5 days period. However the stock moves will also Granger-cause the volume of financial words in tweets. We mentioned above that bullish sentiment will contribute to explain stock returns. But Mao *et al*. (2011) has also shown that the process is circular with stock returns Granger-causing tweets investors' sentiment. The relation is not linearly causal; the polarity of the opinions is not enough to track the informativeness process incorporating information in the stock price. In that sense the buzz hypothesis focuses on the (variation) in attention leading to stock price moves. It is supported by Mao *et al*. (2011).

Results are contradictory. Zheludev (2016) compared the informativeness of both sentiments and volume in leading financial markets prices. He considers tweets sentiments and tweets volume. Sentiments contain higher lead-time information about the futures stock returns in excess of what can be drawn by Twitter messages volume. Divet (2016) used the volume of tweets to compare a sentiment approach rationale with the salience effect of tweets volume. Actually after differentiation he considers variations in tweet volume similarly to Da *et al.* (2011) who look at surge in volume to track a change in the attention from retail investors. The Granger causality is more strongly verified (at the 1% level) for a 5 hours period preceding the stock price move. The variation of volume of tweets featuring a rise in attention will help in explaining the subsequent move in the index up to 5 hours later.

However, empirical test of the buzz hypothesis are rare as this hypothesis is relatively new.

6. Conclusion and perspectives of research

The information integration process still takes some time even if the immediacy of social media and networks speed the integration of news. The immediate chocks of prices as they are instrumented and defined with CARs are not sufficient to cover the information process. The feeding of stock prices may be now measured through hourly sub-periods. Social media defines permanent communication news places. Intraday may be more relevant in analyzing than daily stock prices. The trend to use continuously real time news is an avenue of research that is expected to develop (Li *et al.*, 2015; Divet, 2016).

The continuous real time production and diffusion of pieces of information in the social media are the ultimate step where medium and message are actively produced and used by the end-user of the information. In that sense we are in the McLuhan's situation where the medium is the message. This is new. This is similar to an open- pit where a buzz is created and where information circulates. The content of information (message) is only one aspect of the communication process; the medium is in itself important as it raises attention. The literature has outlined that increase in the volume of tweets will signal attention of the (retail) investor and will raise the transaction volume. The loudness of the buzz has by itself an informational value. Using the conclusion of Sprenger *et a*l. (2014, p792) : "investors discussion in an online forum meaningfully reflects real-world news events". It influences future prices but also future transactions. The process of attention and opinion are linked. The information theory will emphasize the role of the sentiments that are pieces of news information and that contribute to a mood with regard to the firm and its stock value. The salience theory (or buzz theory) will privilege the attention dragging mechanism resulting from a buzz. The variables used to measure these dimensions are opinions and (variation in) volume of messages. Both explanations can co-exist. Some authors will support that sentiments are more important in leading stock prices moves than volume. Others will recognize that the loudness of the buzz is the first driver of future stock price moves.

Both the sentiment/information and the attention/salience approaches will initiate circular consequences. On the short term sentiments will influence stock returns in way of their polarity. On a longer term horizon, the stock price is expected to go back to the previous state. So the relationship between sentiment and the direction of the stock move may be difficult to identify and on the whole can be limited. Attention, referred to as a surge in social media volume, leads stock price moves as outlined in many studies. However this is a Granger-causality which means that buzz is relevant in setting future stock prices. But a reverse causality phenomenon can also be demonstrated. Stock prices moves are new pieces of information that will contribute to increase the volume of information diffused between investors through social media.

These circularities underline a process of diffusion that gives strengths to the concept of informativeness. Noise is a dimensional aspect of this process similarly with transaction volumes. This is why the hypothesis of withe noise that refers to *i.i.d.* residuals in market model regressions should be questioned. It refers to the efficient market hypothesis where the stochastic process of stock returns is affected by new valuable and independent pieces of information that are immediately integrated in the price. Buzz and attention will create noises and interdependencies in the information integration process. In that sense the buzz approach support the Roll's measure of informativeness based on idiosyncratic risk. The R2 measure he advocates relies on the variation through time of risk specific residuals. In that sense, other alternative measures of informativeness based on volume changes are also relevant (Amihud, 2002; Llorente *et al.*, 2002).

The analysis of an event by giving it a polarity i.e. by associating it a positive or a negative opinion is complex and methodologically challenging. The concept of buzz in the financial market allows to identify a circular rationale and integrate the idea that time is an element of the information process. From an empirical point of view the measure of attention is easily achieved by the volume of search queries, by the number of pieces of information, by the number of opinions through tweets without questioning the polarity of the message. It may help to design future empirical researches as, when "the medium is the message", the analysis of the content of the message is no more necessary.

Acknowledgements

I thank A. Coen, A. Divet and L. Xu for their valuable comments and suggestions**.**

References

Amihud Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5: 31-56.

Arditi E, Yechiam E, Zahavi G. 2015. Association between stock market gains and losses and Google searches. *PLoS ONE* 10(10): e0141354.

Barber B, Odean T. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21: 785-818.

Boulland R, Degeorge F, Ginglinger E. 2016. News dissemination and investor attention. *Review of Finance* forthcoming.

Bushee B, Core J, Guay W, Hamm S. 2010. The role of business press as an information intermediary. *Journal of Accounting Research* 48(1): 1-19.

Bushee B, Miller G. 2012. Investor relations, firm visibility, and investor following. *The Accounting Review* 87(3): 867-897.

Da Z, Engelberg J, Gao P. 2011. In search of attention. *Journal of Finance* 66(5): 1461-1499.

DellaVigna S, Pollet JM. 2009. Investor inattention and Friday earnings announcements. *Journal of Finance* 64: 709–749.

Divet A. 2016. Can Twitter predict financial markets? *Master's thesis, Université Paris Dauphine*.

Dubois M, Dumontier P. 2006. « Do Conflicts of Interest Affect Analysts' Forecasts and Recommendations? A Survey », *in* Thévenoz, L., Bahar R. (Eds.), *Conflicts of interest, corporate governance and financial markets*, Kluwer Law International, 187-210.

Go A, Bhayani R, Huang L. 2009. Twitter sentiment classification using distant supervision. *Entropy* 17.

Heston S, Sinha N. 2016. News versus sentiment: Predicting stock returns from news stories. *FED Finance and Economics Discussion Series* 2016-048.

Li E, Shen R, Wu J. 2015. Do analyst stock recommendations piggyback on recent corporate news? An analysis of regular-hour and after-hour revisions. *Journal of Accounting Research* 53(4): 821-861.

Llorente G, Michaely R, Saar G, Wang J. 2002. Dynamic volume-return relation of individual stocks. *Review of Financial Studies* 15: 1005-1047.

Mao H, Counts S, Bollen J. 2011. Predicting financial markets: Comparing survey, news, Twitter and search engine data. *eprint arXiv* 1112.1051.

Michaely R, Womack K. 1999. Conflict of interest and the credibility of underwriter analysts' recommendations. *Review of Financial Studies* 12: 653-686.

Sprenger T, Sandner P, Tumasjan A, Welpe I. 2014. News or noise? Using Twitter to identify and understand company specific news flow. *Journal of Business Finance & Accounting* 41-7&8: 791-830.

Tetlock P. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62: 1139-1168.

Tetlock P, Saar-Tsechansky M, Macskassy S. 2008. More than words: quantifying language to measure firms' fundamentals. *Journal of Finance* 63: 1437-1467.

McLuhan M. 1964. Understanding Media. *Routledge*, London.

Zheludev I. 2016. When can social media lead financial markets?, *PhD. Thesis, University College London.*