Sensitivity to Sentiments: Social vs News Media Impacts on Stock Markets - A comparison using textual analysis *

Baoqing Gan†, Vitali Alexeev, Ron Bird, and Danny Yeung

Finance Discipline Group, UTS Business School, Sydney, Australia

November 11, 2018

Abstract

Relying on information embedded in textual news, researchers are increasingly focusing on using automatically-determined news sentiment to explain stock price movements and alert users of sentiment-based investment opportunities. Using sentiment scores from Thomson Reuters MarketPsych Indices (TRMI) - sentiment measures generated by an algorithm that extracts texts from major social and news media outlets, this paper analyses overall market sensitivity to media sentiments from 2011 to 2017, and compares different effects of social and news media. We find that social media tends to have stronger effect on stock market than news media does after February 2014 - a period where SEC admits the thriving social media as official information dissemination channels for public companies. Further analysis employing Structural Vector Autoregressive (SVAR) model reveals new empirical facts about the dynamic relationships between media sentiments and aggregate stock market activities. Overall, our insights and findings help shed light on how information is incorporated into stock prices in current interactive digital media era.

Keywords: sentiment scoring; textual analysis; big data; vector autoregressive (VAR) model

JEL: G14, G40, G41

*We are grateful to Romain Legrand, Alistair Dieppe, and Björn van Roye from European Central Bank (ECB) for providing MATLAB's **Bayesian Estimation, Analysis and Regression (BEAR) Toolbox**; we thank Thomson Reuters Financial and Risk for offering MarketPsych Indices (**TRMI**) as part of our research data; we are also grateful to comments from assessors and participants at the 2018 UTS Business School Autumn Finance Discipline Group Stage Assessment Presentation.

†PhD student and casual academic at UTS Business School, Email: *baoqing.gan@student.uts.edu.au*; *baoqing.gan@uts.edu.au*; Phone: +61 2 9514 7787. This research includes preliminary results and constitutes a part of my PhD Thesis. I am grateful to the continuous support, guidance and encouragement from my supervisory panel: Dr. Vitali Alexeev *vitali.alexeev@uts.edu.au*, Prof. Ron Bird *ron.bird@uts.edu.au*, and Dr. Danny Yeung *danny.yeung@uts.edu.au*.

1 Introduction

The financial sentiment literature has shown that macroeconomic announcements, major geopolitical events, and corporate announcements change investors' sentiments and often influence stock prices. Traditionally, investors receive this information through mainstream financial news reports, official announcements, corporate conference calls, and analysts research reports. Nowadays, social media outlets such as StockTwits and internet message boards/forums are becoming more prominent in information dissemination process, providing exponentially increasing quantities of company related information to the market, creating attention-grabbing hot topics. These topics sway investors' beliefs about company's future outlook, thus forming investor sentiments that ultimately impact stock prices. The classical asset pricing model assumes that investors interact and mutually influence each other only by market price mechanisms. This assumption simplifies the information dissemination process and overlooks the social effect and interactions between investors. In reality, investors communicate and learn information through a combination of news media and social media, making social influence an essential factor in information dissemination, and thus asset pricing (Hirshleifer and Teoh, 2008). As early as 1896, Le Bon (Le Bon, 1896) pointed out that when people are in certain groups, they will behave quite differently from when they are alone. Group sentiments are contagious. Individual's behaviour varies in accordance with their social contexts. Similarly, Fehr and Tyran (2005) found that a small amount of individual irrationality may lead to large deviations from the aggregate predictions of rational models under certain circumstances. News media, online or paper alike, plays a crucial role as the storyteller and information-transmitter for social interactions, which will, in turn, influence the stock markets dynamics.

In the recent decade, advancements in digital technology and mobile devices make social media like Twitter an increasingly crucial channel for stock information sharing.[1] In early 2013, Bloomberg LP announced that it would add a small number of Twitter accounts to its financial information terminals, which are commonly used by traders on Wall Street.[2] Just a few month later, on 23 April 2013, a fake tweet from official Twitter account of the Associate Press announced that President Obama was injured in two explosions in the White House.[3] According to Washington Post, this Syrian hacked tweet was retweeted 4,000 times in less than five minutes with its nearly 2 million followers. Dow Jones Industrial Average (DJIA) dropped 143.5 points within 2 minutes; S&P 500 temporarily lost an estimated US$136 billion in value. Some claim that the financial industry might rely too heavily on the trading algorithms based on social media content.

The US Securities and Exchange Commission (SEC) also reacts to the modern developments

[1]Stafford, P. (2015), 'Traders and investors use Twitter to get ahead of market moves', *FINANCIAL TIMES*, April 29, accessed 12 August 2018, <https://www.google.com.au/amp/s/amp.ft.com/content/c464d944-ee75-11e4-98f9-00144feab7de>.

[2]Alden, W. (2013), 'Twitter arrives on Wall Street, via Bloomberg', *The New York Times*, April 4, accessed 12 August 2018, <https://dealbook.nytimes.com/2013/04/04/twitter-arrives-on-wall-street-via-bloomberg/>.

[3]Fisher, M. (2013), 'Syrian hackers claim AP hack that tipped stock market by $136 billion. Is it terrorism?' *The Washington Post*, 23 April, accessed 12 August 2018, <https://www.washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/?utm_term=.5e2044c627e4>.

in corporate information dissemination. After issuing a guidance in 2008 admitting that websites can serve as an effective means for disseminating information to investors, the SEC pointed out in April 2013 in its investigation report toward Netflix that "company communications made through social media channels could constitute selective disclosures and, therefore, require careful Regulation Fair Disclosure (Reg FD) analysis". This investigation report stems from the CEO of Netflix posting on his Facebook account that Netflix's monthly online viewing had exceeded one billion hours for the first time, but without disclosing this information through Form 8-K or other press releases. As a result, Netflix's stock price increased from \$70.45 at the time of the Facebook post to \$81.72 at the close of the following trading day.[4] As a continuation of the SEC's warming up to social media, which began in April 2013 when it approved the use of postings on Facebook and Twitter to communicate corporate announcements, SEC's staff said in a "Compliance and Disclosure Interpretations" in June of 2015 that a start-up firm can post Twitter message about its stock or debt offering to gauge interest among potential investors (Bartov et al., 2018).

The above examples and new SEC publication requirements demonstrate that stock markets are sensitive to news and social media information alike. They indicate that financial industry is recognising the importance of new information channels. Major **event** occurs, which is covered by news and social media that catches investors' **attention**, forming new **beliefs** about future cash flows, and as a result, generating group **sentiment** that finally impacts on **market variations**. However, the precise mechanics along this chain remains unclear. More importantly, the functional difference between news and social media in this process is rarely studied, either.

In this paper, we aim to address the following questions: How the financial news landscape changed in the recent decade? In different time period, how different types of media (i.e. social and news media) impact on the aggregated stock market? How social and news media sentiments react to stock market shocks? In general, answers to these questions help reveal the dynamic associations between media activity (both quantities and emotions) and market variations, generate new insights about the information dissemination process - one of the most important topics in modern finance, and contribute to literature on how information is incorporated into stock prices.

We first provide detailed statistical analysis on social and news media sentiment indices of major US stock indexes (DJIA and S&P 500). To the best of our knowledge, there are only a few studies[5] that slice and dice such kind of research data from the perspective of contrasting social and news media. We then investigate the interplay between the two main series of media activeness measures. We find that after February 2014, social media activeness began to significantly lead news media, suggesting that social media tends to generate more influences and that impacts from news media are decreasing. This phenomenon may well be due to the aforementioned SEC's new announcement policy. We continue our study by systematically analysing the dynamic relationships between S&P 500 market activity variables (return, volume and volatility) and media

[4] The US Securities and Exchange Commission 2013, *SEC Says Social Media OK for Company Announcements if Investors Are Alerted*, Press Release, accessed 12 August 2018, <https://www.sec.gov/news/press-release/2013-2013-51htm>

[5] See Shen et al. (2017), Audrino and Tetereva (2017), and Huang et al. (2018)

activity variables. Consistent with Jiao et al. (2016a), we find significant heterogeneities in market reactions to news and social media innovations, while feedback effects from market shocks to both social and news media are homogeneous. Our findings that return helps predict social media sentiment but social media sentiment does not predict future returns are in line with Brown and Cliff (2004). Our results that news media sentiment and S&P 500 short-term return mutually cause each other, and that negative news media sentiment also cause S&P 500 trading volume are in accordance with Tetlock (2007). Finally, comparing subsamples before and after February 2014 - a critical point we identified in our first finding, help elaborate furthre details of how social media effects turn stronger and news media effects become weaker in recent years.

The rest of the paper is organized as follow: Section 2 reviews past works, Section 3 describes sample data and discusses research methodology, Section 4 reports results, and Section 5 concludes and points out our future research agenda.

2 Literature Review

2.1 Sentiment and Market

Investor sentiment is defined as the net amount of any group of investors' optimism or pessimism, reflected in any asset or market price at a specific time (Kirkpatrick II and Dahlquist, 2010). It is the collective emotional and other psychological factors that come from human interaction involved in determining price deviations. Another broad definition of investor sentiment states that it is a belief about future cash flows and investment risks that is not justified by the facts at hand (Baker and Wurgler, 2007). The behavioural model by De Long et al. (1990) argues that the difficulty to predict irrational investors' sentiment changes creates a risk in asset price. This sentiment risk deters rational investors from betting against the irrationals because of the trading costs, and as a result causes stock prices to deviate from their fundamental values even without fundamental risk. The implications of De Long et al. (1990) theory is further explored by numerous empirical researches on stock market anomalies such as closed-end fund puzzle, equity premium puzzle, and excess volatility of asset prices. A leading exploration is from Lee et al. (1991), who verify that closed-end fund discount (CEFD[6] hence force) can be used as a proxy for irrational investors' negative sentiment. Lee et al. (1991) further suggest that movements in security prices (return) may be attributed to fluctuations in investor sentiment. Other theoretical models addressing the role of investor sentiment on market behaviours include: Grossman and Stiglitz (1980), Black (1986), Campbell and Kyle (1993), Barberis et al. (1998), Daniel et al. (1998), and Hong and Stein (1999).

Based on these behavioural hypothesis, a plethora of ideas have been proposed. They aim at answering the questions of how to measure investor sentiment more accurately, and how to test return predictability from sentiment more effectively. For instance, Neal and Wheatley (1998) compare return predictability from three popular individual investor sentiment measures: the CEFD (Zweig, 1973), the odd-lot sales to purchases ratio (Hardy, 1939), and the net mu-

[6] A full list of all acronyms used in this study and their explainations is provided in the appendix Table A.1.

tual fund redemptions (Swaminathan, 1996). They find that the CEFD and net redemptions predict size premium, but odd-lot ratio does not help forecast return. Qiu and Welch (2004) compare investor sentiment measures constructed from UBS/Gallup sentiment survey and the CEFD sentiment proxy, and find that consumer confidence measures might be better proxies for investor sentiment. The most famous sentiment measures using fundamental market variables are proposed by Baker and Wurgler (2007). Using principal component analysis (PCA), they consider six major proxies: trading volume (NYSE turnover), the dividend premium, closed-end fund discount, number of IPOs, first-day return of IPOs, and new equity issues. Their sentiment indexes capture a prevailing "greed" versus "fear" or "bullish" versus "bearish" notion generally. They find that small, young, unprofitable, high-volatility, non–dividend paying, growth stocks, or stocks of financially distressed firms may be more sensitive to big investor sentiment shocks. Similarly, Brown and Cliff (2004) use Kalman filter to generate sentiment indicator that combines direct sentiment measures (e.g. Survey sentiment from American Associate of Individual Investors) and indirect sentiment measures (the Baker and Wurgler sentiment measures). In a VAR system comprising of market variables and sentiment indicators, they find that sentiment is not limited to individual investors, that returns and contemporaneous sentiment are strongly positively related, and that returns predict future sentiment, but sentiment does not predict future returns.

2.2 Textual Analysis and Sentiment Measures

A shortcoming of both market fundamental data based and survey based sentiment measures is that it only suits to the aggregate market rather than individual stocks. To overcome this limitation, sentiment measures using new data sources and methods are also proactively created. Textual analysis from stock-information related documents is one of the most prominent approach to measure both market level and company level tonality. Moreover, due to the ubiquity of social media platforms as information diffusion channel, the interactive user-generated information offers new opportunities to study investor sentiment expressed within these digital platforms. Textual analysis is the tool that help academics to realize these goals.

Textual analysis includes the process of translating qualitative text into quantitative measures and measure the degree of positivity and negativity in texts (Nardo et al., 2016). According to a survey from Nardo et al. (2016), lexical analysis algorithms select fit-to-the-purpose words and phrases, and then determine their positive/negative tonality or more complex feelings (emotions) based on some externally provided dictionaries or semantic rules. Top four so called "bag-of-words" sentiment classification dictionaries are: Harvard General Inquiries (GI), Henry (2008), Diction, and Loughran and McDonald (2011b) (LM henceforce). LM is proved to be most appropriate in accounting and finance applications because it provides a relatively more exhaustive word list within the context of modern business lexical environment. Loughran and McDonald (2011b) analyse texts in a large sample of 10-K filings from 1994 to 2008, and they show that nearly three-fourths of words identified as negative by the common sociopsychological Harvard Dictionary were not considered negative in financial contexts. They develop a new word list that is more relevant to topics of returns, trading volume, volatility, fraud, corporate fundamental weakness, and earnings surprises. Yet as the quantity of information generated in news and social media increasing explosively, it is difficult to quantify the content of interest only by traditional methods such as word counting and categorization. Thus, computational linguistic algorithms such as Naïve Bayes and Support Vector Machine are more suitable in the age of big data.[7]

2.3 Previous Empirical Research in Media Text Sentiment

Depending on information source, empiricists investigate qualitative content in 1). corporate disclosures (e.g. SEC filings, press release, and conference call scripts), 2). professional news articles, 3). internet message boards, and 4). other kinds of social media such as Twitter and StockTwits. Kearney and Liu (2014) use these four categories of data source and summarize research papers up to 2013 in a survey paper. However, after 2013 more empirical literature now emphases on Internet expressed sentiment and uses algorithmic textual analysis on mixed source (Kearney and Liu, 2014).

The first strand of literature analyses corporate regulatory filings such as 10-K or 10-Q. Considering both positive and negative tones in firm 10-K filing, Jegadeesh and Wu (2013) generate a term weighting scheme that calculate frequency of positive or negative words relative to total document word length. They compare their term-weighting sentiment measure with LM's weighting method and find that market underreact to 10-K tonality during the filing period, but similar underreaction is not reflected using LM's inverse document frequency weights method. Jegadeesh and Wu (2013) point out this discrepancy and indicate that the term weighting scheme is more important than the accuracy and completeness of underlying lexicons. In another empirical work of Loughran and McDonald (2011a), the impact from accounting "red flag words" in corporate 10-K filings are analysed. Loughran and McDonald (2011a) investigate the relationships between these "red flag words" and filing period return, post filing period return volatility, analyst forecast dispersion, and fraud litigation likelihood. They find that terms like *unbilled receivables* signal a firm may be accused of fraud subsequently, while phrases such as *substantial doubt* are significantly related to filing date negative excess returns, higher volatility, and greater analyst earnings forecast dispersion. Their findings suggest that corporate governance wording signals conveyed in management filings may help explain the filing day drift.

Another strand of literature uses professional news articles as textual analysis data source. Some of the most highly mentioned researches include: Tetlock (2007), Tetlock et al. (2008), Engelberg (2008) and Garcia (2013). Tetlock (2007) uses General Inquirer (GI) dictionary as a reference to count negative words on the Wall Street Journal's "Abreast the Market" column from 1984 to 1999. He constructs a pessimistic indicator from 77 categories of negative words from principal component analysis and builds up a 5-lag Vector Autoregressive (VAR(5)) model to investigate the lead-lag relationships among the negative sentiment measure, the Dow Jones Industrial Average (DJIA) returns, and the New York Stock Exchange (NYSE) trading volumes. He finds that negative sentiment and short-term DJIA returns mutually cause each other, and that neg-

[7] Loughran and McDonald (2016) comment the advantages and disadvantages of Naïve Bayes Method.

ative sentiment also causes NYSE trading volumes. After detecting the market-level negative sentiment impact on aggregated stock returns and volumes, Tetlock et al. (2008) continue to explore media sentiment effects at firm level using fraction of negative words method, and they find that high fractions of pessimistic word about corporate fundamentals in news texts predict low firm earnings. This research suggests that lexical content in news media is capable of capturing the ambiguous corporate fundamental characteristics, which will quickly incorporate into stock prices. To overcome the shortcoming of negative only characteristic of sentiment measure in Tetlock (2007) and Tetlock et al. (2008), Garcia (2013) constructs both positive and negative fraction of words from financial news on New York Times columns. He finds that news content assists to forecast stock returns at daily frequency, but only during recessions. The benefit of Garcia (2013) lies in its linkage of news media sentiment with 1905-1958 major business recessions. This research also uses VAR model to investigate the feedback effect from stock return to news sentiment. But Garcia (2013) only accounts for sentiment effects on returns, other market feature variables are not investigated.

Studies that rely information source on internet message boards are exemplified by Antweiler and Frank (2004) and Chen et al. (2014). A pioneering research in this realm is proposed by Wysocki (1998). It proves that *Yahoo!Finance* posting volumes could be used to predict trading volume on the next day. Sparkled by this research finding, Antweiler and Frank (2004) apply computational linguistics to analyse the tonality of message postings on *Yahoo!Finance* and *Rating Bull* related to 45 DJIA companies, controlling for Wall Street Journal news stories effects. Antweiler and Frank (2004) analyse the correlation between message board variables (messages, words, bullishness, and agreement) and stock feature variables (return, volatility, volume, and bid-ask spread), and run contemporaneous and time-sequencing regressions on these variables subcequently. They find that increases in message board posting volumes predict decreases in return on the next day. Also, both the magnitude of disagreement among these postings and the quantities of postings could predict trading volume: the higher the disagreement, the lower the trading volume on the next day. Chen et al. (2014) examine the postings on investment discussion forum *seeking alpha* and find that opinions conveyed in both posting articles and commentaries help predict stock returns and earnings surprises. They use negative words fractions in message board posting articles and commentaries as main sentiment proxies, controlling for news media effect (DJNS/WSJ coverage and tonality) and analyst advice. Main findings from this research demonstrates the predictability of user-generated information on social media platforms to stock returns and earnings.

Researches that extract information from other social media platforms like Twitter and Facebook are represented by Sprenger et al. (2014), Siganos et al. (2014) and Renault (2017). This strand of literature extends financial information dissemination to the social influence and interactions, which may be more informative than traditionally single-direction information publication process. Following Antweiler and Frank (2004) framework, Sprenger et al. (2014) analyse company-level Tweets of S&P 100 companies. Their exploration between Tweets' features (bullishness, message volume, and agreement) and market features (return, traded volume and volatility) use contemporaneous regression and Fama-MacBeth regression to show that feedback effect from market feature variables to Twitter feature variables are stronger than from Twitter to stock market. They also find that bullishness (fraction of buy over sell) in tweets relates to level of return but not abnormal return. The combination of these evidences indicates that social media platforms contain stock value related information, but this information may not be fully aggregated into the stock market contemporaneously. Similarly, Siganos et al. (2014) use daily sentiment index from Facebook's Gross National Happiness to show that sentiment is contemporaneously positively correlated with stock return, and that negative sentiments are related to spikes of trading volume and stock volatility. Siganos et al. (2017) further point out that divergence of sentiment (disagreement) impacts on stock market, which goes beyond the effect from level of sentiment. They provide evidence that high volume of polarized sentiment is positively related to trading volume based on Facebook's happiness index, which contradicts the evidence from Antweiler and Frank (2004), who suggest higher disagreement predict lower trading volume. Renault (2017) construct their proprietary algorithm that abstract sentiment from StockTwits and offer evidence that online investor sentiment helps predict stock index return at intraday level. They also prove that this intraday sentiment effect is due to novice traders.

Another branch of empirical researches (Preis et al., 2013; Liu and McConnell, 2013; Ranco et al., 2015; Jiao et al., 2016b) have shown that both the volume of reports/postings and the attitudes conveyed within them are influential to stock market and corporate events. Liu and McConnell (2013) show that both the media attention level (measured by number of articles in proposed acquisition) and the tone within these articles (measured by the percentage of negative words using LM dictionary) are significantly associated with the success of acquisition deal. Jiao et al. (2016a) is probably the closest research to this paper because it distinguishes different roles news and social media information play at both market level and individual stock level. Using panel VAR estimation, they find that increases in news media activities have positive and significant impact on subsequent social media activities, but influences from rising in social media activeness on news media activeness is insignificant. Besides, their stock level and market level empirical results show that high social media reports predict high return volatility and high trading activity, whilst intense news media coverage forecasts low volatility and low trading activity.

And there is another kind of study that concentrates only on active posting volumes like media coverage rate and internet searching volume, to proxy for investors' attention so as to quantify the inter-relationships between investor attention and market reactions. For instance, Fang and Peress (2009) use portfolios of all NYSE and 500 NASDAQ companies, and coverage of these companies on four major newspapers: New York Times, USA Today, the Wall Street Journal, and Washington Post to study the relationship between media coverage and market reaction from 1993 to 2002. Controlling for major risk factors, they find that stocks not covered by media earn significantly higher future returns than stocks that are heavily covered. They assert that media effect could be explained by Merton (1987) Investor Recognition Hypothesis - that thinly covered stocks should pay a premium to investors. However, what contradicts to this finding is provided by Heston and Sinha (2017), who use Thomson Reuters Neural Network sentiment scores to show that companies without news coverage underperform companies with news at weekly frequency. One problem of using media coverage to measure investor attention lies in the gap between media coverage rate and investor reading rate, because attention is a scarce resource nowadays. Da et al. (2011) use Google Search Volume (GSV) Index of Russell 3000 index from *Google Trend* to measure investor attention, and they argue that searching volume is a more "active" measure for attention than media report volume. They claim that Google search volume could be used to proxy for retail investor's attention in a timely fashion. They find that increases in GSV predicts higher stock prices fortnightly, and this upward stock price drift reverse to normal within one year. Their results help explain decade-old anomalies like IPO first day excess return and future reversion from a different angle.

2.4 Motivation and Contribution

There are several shortcomings within the previously mentioned researches. Firstly, although researches that focus only on new information channel have been developed for a large extent since 2013, limited attention has been paid to the discrepancies between social media and news media sentiment effects. Secondly, relative smaller numbers of empirical studies taking both volume of media activity and emotions expressed within them into account. Thirdly, some researches using linear regression models often neglect the dynamic relationships (feedback effect) and cross-dependencies between market variables and sentiment fluctuations. Fourthly, most of the textual sentiment measures are binary (positive/negative) or categorical (positive/neutral/negative) by construction, it is less common to observe relatively more accurate absolute/net levels of sentiment measures.

This paper aims to narrow these literature gaps and contribute to this topic in the following three ways. To begin with, we focus on contrasting different effects social and news media functions on aggregate stock markets, and explaining how the information environment and financial news landscape have changed. Meanwhile, allowing for lead-lag relationships and incorporating both media volume (*Buzz*) and emotional (*Sentiment*) measures, we test heterogeneous impacts of social and news information on stock market at different subsampling periods. Last but not least, using net levels of sentiment proxies (i.e. positive minus negative sentiment), rather than categorical sentiment scores, we account for dynamic relationships between market variables and sentiments within social and news media. In general, this paper reveals new facts of past and present in a different perspective and will shed light on how information is incorporated into stock prices in the new digital information era.

3 Data and Methodology

Our dataset is comprised of two sources: sentiment data and stock market data. In this section, we provide details on each of the datasets and describe our data pre-processing methods.

3.1 Sentiment Data

TRMI incorporate analysis of news and social media in real-time by converting the quantity and variety of financial economic news and internet messages into manageable information flows.[8] It provides three content types: **news**, **social** and **combined**. We use 24-hour observation window company groups (market index equivalent) data from 2011 to 2017. These daily data are updated at 3:30pm US Eastern time each day, including weekends and other non-trading days. Company group sentiment is aggregated scores for the largest stocks by market capitalization. There are two main company groups for the US market: the top 500 (*MPTRXUS500*) which aims at capturing S&P 500 sentiment, and the top 30 (*MPTRXUS30*) that targets at capturing DJIA sentiment.

TRMI offers three main types of sentiment indicators: 1) **Emotional** indicators including *Anger*, *Fear* and *Joy*; 2) **Macroeconomic** metrics such as *Long vs Short*, *Earnings Forecast*, and *Interest Rate Forecast*; and 3) **Buzz** metric, a topic heat measure on market-moving news such as *Litigation*, *Mergers*, and *Volatility* for a specific company or company group. Table 1 and Table 2 summarize descriptive statistics of S&P 500 Social and News sentiment indices. Summary statistics for DJIA corresponding TRMI Social and News indices are provided in Table A.4 and Table A.5 in the appendix. Panel (A) of these four tables are polarized scores that ranges [-1,1], and panel (B) of these four tables are unidirectional scores which is in the realm of [0,1]. All polarized sentiment scores are 24-hour rolling averages of its respective overall net references (positive references net of negative references). For example, at a specific 24-hour window, sentiment score of S&P 500 index equals -0.2, which means all negative sentiment texts referring to S&P 500 index net of all positive sentiment texts have a value of -0.2 as calculated by TRMI. From these descriptive tables, we observe three direct facts: firstly, *Buzz*, a sheer media coverage volume metric for both social and news media, has a much more larger absolute value than other emotional proxies. Secondly, other emotional sentiment measures e.g., *optimism*, *joy*, and *anger* are scaled close to zero by their linguistic scoring methods. Thirdly, we find more empty values in news sentiment scores than for social media, probably resulting from the fact that news reports require more stringent censorship procedures than social media. Lastly, both Durbin-Watson (DW) test and Ljung-Box test to 5 lags (LB-5) show evidence of autocorrelation with potential long memories for all available social and news sentiment indices.

[8] The data are provided by Thomson Reuters Financial and Risk Team as part of TRMI product. *Thomson Reuters MarketPsych Indices 2.2 User Guide*, 23 March 2016, Document Version 1.0.

Table 1: Descriptive Statistics for TRMI MPTRXUS500 Company Groups Social Indices. Sample period 01/Jan/2011 - 30/Nov/2017; sentiment indices are grouped into polarized scores that ranges [-1,1] and unidirectional scores that bounded at [0,1]. Buzz is a special measure which has different scale from all other metrics. Data in *laborDispute* was too sparse during our sample period, we include it here for completeness. Durbin-Watson test and Ljung-Box 5 lags test for all indices show there is autocorrelation.

	Mean	Std	Max	Min	Skew	Kurt	25th	Median	75th	ITQ
Panel (A): Polarized Groups [-1,1]										
sentiment	-0.020	0.030	0.082	-0.127	-0.32	2.80	-0.040	-0.016	0.001	0.042
optimism	0.000	0.008	0.020	-0.034	-0.40	3.11	-0.005	0.001	0.005	0.010
loveHate	0.006	0.002	0.023	0.000	3.17	21.58	0.005	0.006	0.006	0.001
trust	-0.001	0.002	0.016	-0.021	-0.97	15.12	-0.003	-0.001	0.000	0.002
conflict	0.020	0.005	0.081	-0.002	2.70	21.92	0.017	0.020	0.023	0.005
timeUrgency	0.019	0.004	0.049	0.004	0.70	5.76	0.016	0.019	0.021	0.005
emotionVsFact	0.531	0.023	0.627	0.407	-0.20	4.54	0.518	0.532	0.546	0.029
marketRisk	-0.008	0.004	0.023	-0.027	-0.19	5.03	-0.011	-0.008	-0.005	0.005
longShort	0.004	0.004	0.090	-0.039	7.08	163.95	0.002	0.004	0.005	0.004
longShortForecast	0.001	0.001	0.003	-0.008	-1.87	24.97	0.000	0.001	0.001	0.001
priceDirection	0.003	0.002	0.014	-0.007	-0.04	4.33	0.002	0.003	0.004	0.003
priceForecast	0.001	0.000	0.003	-0.001	0.14	5.35	0.000	0.001	0.001	0.001
analystRating	0.001	0.001	0.008	-0.006	0.56	12.05	0.000	0.001	0.001	0.001
dividends	0.001	0.001	0.008	-0.004	2.12	25.00	0.001	0.001	0.001	0.001
earningsForecast	0.002	0.001	0.007	-0.003	0.86	6.03	0.001	0.002	0.002	0.001
fundamentalStrength	0.005	0.003	0.018	-0.004	0.86	4.73	0.004	0.005	0.007	0.003
managementChange	0.002	0.002	0.064	0.000	21.32	667.17	0.001	0.002	0.002	0.001
managementTrust	-0.001	0.002	0.016	-0.047	-7.58	114.09	-0.001	0.000	0.000	0.002
Panel (B): Unidirectional Groups [0,1]										
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
anger	0.014	0.003	0.041	0.007	1.61	11.83	0.012	0.013	0.016	0.004
fear	0.005	0.001	0.010	0.003	0.98	6.86	0.005	0.005	0.005	0.001
joy	0.015	0.002	0.028	0.008	1.02	5.01	0.013	0.015	0.016	0.003
gloom	0.028	0.004	0.056	0.018	0.80	5.10	0.026	0.028	0.031	0.005
stress	0.056	0.004	0.099	0.044	1.35	15.43	0.054	0.056	0.058	0.004
surprise	0.008	0.001	0.026	0.005	2.23	21.96	0.007	0.008	0.009	0.002
uncertainty	0.023	0.003	0.035	0.012	-0.02	3.65	0.021	0.023	0.024	0.003
violence	0.029	0.005	0.063	0.021	1.90	8.72	0.026	0.028	0.031	0.005
volatility	0.026	0.003	0.055	0.019	1.47	10.56	0.024	0.026	0.028	0.004
debtDefault	0.004	0.001	0.018	0.002	2.07	15.73	0.003	0.004	0.005	0.001
innovation	0.003	0.001	0.011	0.001	1.02	6.48	0.002	0.003	0.003	0.001
laborDispute	-	-	-	-	-	-	-	-	-	-
layoffs	0.001	0.001	0.010	0.000	5.63	55.47	0.001	0.001	0.001	0.000
litigation	0.006	0.002	0.024	0.003	2.28	14.89	0.005	0.006	0.007	0.002
mergers	0.004	0.002	0.024	0.001	3.14	22.86	0.003	0.003	0.004	0.002
cyberCrime	0.001	0.001	0.015	0.000	5.53	47.44	0.000	0.001	0.001	0.001
Panel (C): Buzz										
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
buzz	116,484.46	35,769.47	311,543.00	14,179.10	1.37	6.32	94,587.05	110,860.86	130,317.27	35,730.22

Table 2: Descriptive Statistics for TRMI MPTRXUS500 Company Groups News Indices. Sample period 01/Jan/2011 - 30/Nov/2017; sentiment indices are grouped into polarized scores that ranges [-1,1] and unidirectional scores that bounded at [0,1]. Buzz is a special measure which has different scale from all other metrics. Data in *priceForecast, dividends, managementChange, laborDispute, layoffs, and cyberCrime* were too sparse during our sample period, we include them here for completeness. Durbin-Watson test and Ljung-Box 5 lags test for all indices show there is autocorrelation.

	Panel (A): Polarized Groups [-1,1]									
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
sentiment	-0.017	0.037	0.126	-0.173	-0.29	3.22	-0.042	-0.015	0.009	0.051
optimism	0.006	0.007	0.038	-0.037	-0.35	4.39	0.001	0.006	0.010	0.009
loveHate	0.005	0.001	0.013	0.000	0.69	7.18	0.004	0.005	0.005	0.001
trust	-0.001	0.002	0.006	-0.012	-0.86	5.49	-0.002	-0.001	0.000	0.002
conflict	0.032	0.006	0.056	0.017	0.87	4.07	0.028	0.031	0.035	0.007
timeUrgency	0.024	0.004	0.046	0.000	0.06	4.88	0.021	0.024	0.026	0.005
emotionVsFact	0.537	0.028	0.612	0.346	-0.68	4.40	0.521	0.539	0.557	0.036
marketRisk	-0.007	0.004	0.010	-0.031	-0.43	3.84	-0.010	-0.007	-0.004	0.005
longShort	0.002	0.003	0.014	-0.009	0.01	5.17	0.001	0.002	0.004	0.003
longShortForecast	0.000	0.001	0.003	-0.003	0.09	5.67	0.000	0.000	0.001	0.001
priceDirection	0.004	0.003	0.016	-0.012	-0.20	4.28	0.003	0.004	0.006	0.003
priceForecast	-	-	-	-	-	-	-	-	-	-
analystRating	0.001	0.001	0.007	-0.009	-2.16	21.26	0.000	0.001	0.001	0.001
dividends	-	-	-	-	-	-	-	-	-	-
earningsForecast	0.002	0.001	0.008	-0.004	0.60	4.56	0.001	0.002	0.003	0.002
fundamentalStrength	0.008	0.005	0.038	-0.005	1.48	7.35	0.005	0.007	0.010	0.005
managementChange	-	-	-	-	-	-	-	-	-	-
managementTrust	0.001	0.003	0.019	-0.017	-1.11	9.60	0.000	0.001	0.003	0.003
	Panel (B): Unidirectional Groups [0,1]									
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
anger	0.009	0.002	0.022	0.006	1.87	8.82	0.008	0.008	0.009	0.002
fear	0.007	0.001	0.014	0.004	1.19	6.56	0.006	0.006	0.007	0.001
joy	0.008	0.001	0.015	0.003	0.41	4.21	0.007	0.008	0.009	0.002
gloom	0.023	0.003	0.044	0.016	1.17	7.08	0.021	0.023	0.024	0.003
stress	0.056	0.005	0.078	0.042	0.58	4.09	0.053	0.055	0.059	0.006
surprise	0.007	0.001	0.020	0.004	2.15	17.83	0.006	0.006	0.007	0.001
uncertainty	0.019	0.002	0.030	0.012	0.43	3.52	0.017	0.019	0.021	0.003
violence	0.043	0.010	0.176	0.024	3.10	28.76	0.037	0.041	0.046	0.010
volatility	0.032	0.003	0.060	0.024	1.18	9.66	0.030	0.032	0.034	0.003
debtDefault	0.004	0.001	0.013	0.002	1.76	8.82	0.003	0.004	0.005	0.001
innovation	0.006	0.001	0.021	0.001	1.28	13.98	0.005	0.006	0.007	0.002
laborDispute	-	-	-	-	-	-	-	-	-	-
layoffs	-	-	-	-	-	-	-	-	-	-
litigation	0.011	0.003	0.038	0.005	1.60	9.33	0.009	0.010	0.013	0.004
mergers	0.005	0.002	0.022	0.001	1.68	9.49	0.004	0.005	0.006	0.002
cyberCrime	-	-	-	-	-	-	-	-	-	-
	Panel (C): Buzz									
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
buzz	202,401.31	47,847.27	387,635.55	1,468.90	-0.01	3.91	172,081.500	202,994.290	231,451.110	59,369.610

3.2 Stock Market Data

We choose our sampling period from 01/Jan/2011 to 30/Nov/2017 at daily frequency to avoid the 2008-2010 global financial crisis (GFC) turmoil, but at the same time contain the explosive development phase of internet information and social media. Following Antweiler and Frank (2004), and Sprenger et al. (2014), who employed stock return, volatility, and volume as main stock market activity variables to investigate the relationship between media sentiment and stock market, we choose the following market variable data and display their sources:[9]

- S&P 500 and DJIA indices are obtained from SIRCA
- Volume of market indexes come from Datastream Index ETF daily trading volume
- Volatility is calculated as Realized Volatility using previous one month's data
- VIX for S&P 500 and DJIA Index futures are from WRDS CBOE Index

Table 3 summarizes descriptive statistics of S&P 500 Index statistics. Summary statistics for DJIA is contained in Table A.6 in the appendix.

Table 3: Descriptive Statistics for S&P 500 Index. Sample period 01/Jan/2011 - 30/Nov/2017; return is annualized by multiplying daily values by 252; volume is scaled at 10^5; Durbin-Watson test and Ljung-Box 5 lags test for all indices show there is autocorrelation.

	Mean	Std	Max	Min	Skew	Kurt	25th	Median	75th	ITQ
Return	0.09	1.99	10.42	-15.52	-0.54	8.78	-0.68	0.06	1.07	1.75
Volume	1.27	0.70	7.18	0.28	2.49	14.86	0.81	1.11	1.54	0.73
VIX	16.34	5.58	48	9.14	2.07	8.34	12.85	14.89	17.96	5.11
Volatility	22.89	11.94	79.51	5.47	1.98	8.41	15.19	20.45	26.89	11.7

3.3 Data Aggregation Process

Figure 1 panel (a) and panel (b) plot the autocorrelation functions (ACF) for S&P 500 news and social *Buzz* up to 40 lags, where we observe strong weekly seasonality in raw *Buzz* metrics. We deal with this weekly effect for both news and social media first, and then merge these seasonality adjusted sentiment indices with other market data for all trading days only. When merging the non-trading days (public holidays and weekends) values for sentiment indices, we take **average** value of sentiment indices during these days. For example, sentiment indices on Monday represent averages from Saturday, Sunday and Monday sentiment scores. After combining with stock market data, our sample size decrease from 2,526 observations to 1,803 for each time-series. Panels (c) and (d) of Figure 1 show that weekly pattern of *Buzz* metrics after adjusting and aggregation have significantly decreased, and both $Buzz_S$ (panel (c)) and $Buzz_N$ (panel (d)) are positive stationary autocorrelation process with long memory.

3.4 Correlations

We report the pairwise contemporaneous correlations among all available sentiment indices and market variables in Figure 2. To aid interpretation and comparison of a large number of coef-

[9] A full list of all data source is available in the appendix Table A.2.

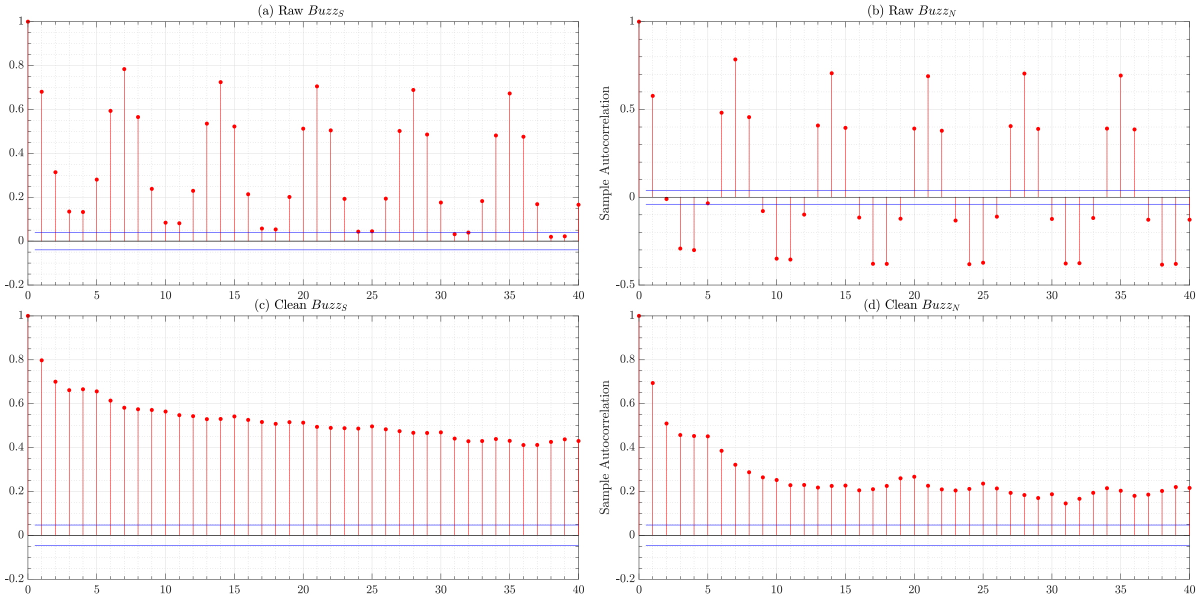


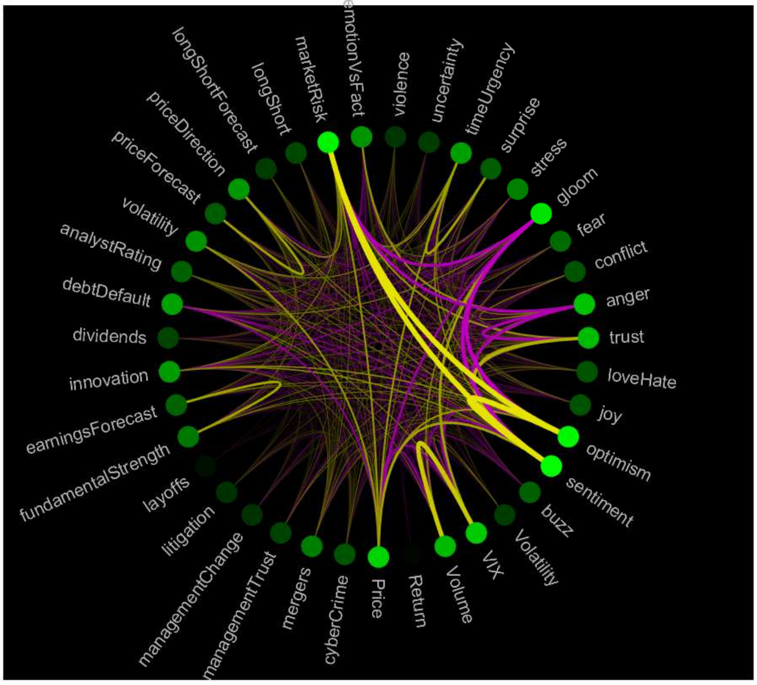
Figure 1: AUTOCORRELATION FUNCTIONS FOR S&P 500 BUZZ SERIES. Panels (a) and (b) show autocorrelation functions for raw social buzz and news buzz series, respectively; panels (c) and (d) are autocorrelation functions of S&P 500 social and news buzz series after adjusting for weekly seasonality and non-trading days. Buzz indices are averaged among non-trading days including any public holidays and weekends.

ficients, we depict correlations in a Schema ball instead of correlation table. Panel (a) and (b) depict associations among social and news indices, respectively. There are less news series because news sentiment indices have more missing values. Yellow curves show positive correlations, and purple lines represent negative correlations. The thickness and brightness indicate strength of correlation relationship, i.e. the thicker the curve, the closer the correlation coefficient is to ±1. We find that social media *Sentiment* and *Optimism* are positively correlated with stock prices, but this relationship is weaker in news media panel. In both social and news panels, both *Sentiment* and *Optimism* are strongly positive correlated with *marketRisk* - a measure defined by TRMI as "bubble-o-meter": the speculative extent relative to rationality.

The correlation estimates in Figure 2 figures are **within group** correlations. Next, we continue generating Kendall correlation plots to investigate **cross group** correlations, that is, the correlation between social and news indices, for six major sentiments: *Buzz, Sentiment, Optimism, Gloom, Fear,* and *marketRisk*. Figure 3 plot this contemporaneous cross group correlations across time. The graph time window is shorter than our sample period because we used first 500 days as estimation window. We find that all six series of our interests are positively correlated with each other for social and news indices. *Buzz* correlation between social and news has the largest range, with highest correlation coefficient amounts to 0.64, whilst the lowest coefficient equals to 0.12. Other correlation pairs are relatively stable across time compared with *Buzz*, but they all exhibit strong heterogeneity across time, suggesting certain noises may exist in these two media, which drives us to further study the time-varying window dynamic relationships between social and news *Buzz*.

In order to study the lead-lag relationships between social and news media, we plot one day lag cross-correlation graphs in Figure A.1 in the appendix. Panel (a) is the correlation between yesterday's social TRMI indices with today's news TRMI indices, while panel (b) is the

(a) MPTRXUS500 Contemporaneous Correlation (2011-2017) Social



(b) MPTRXUS500 Contemporaneous Correlation (2011-2017) News

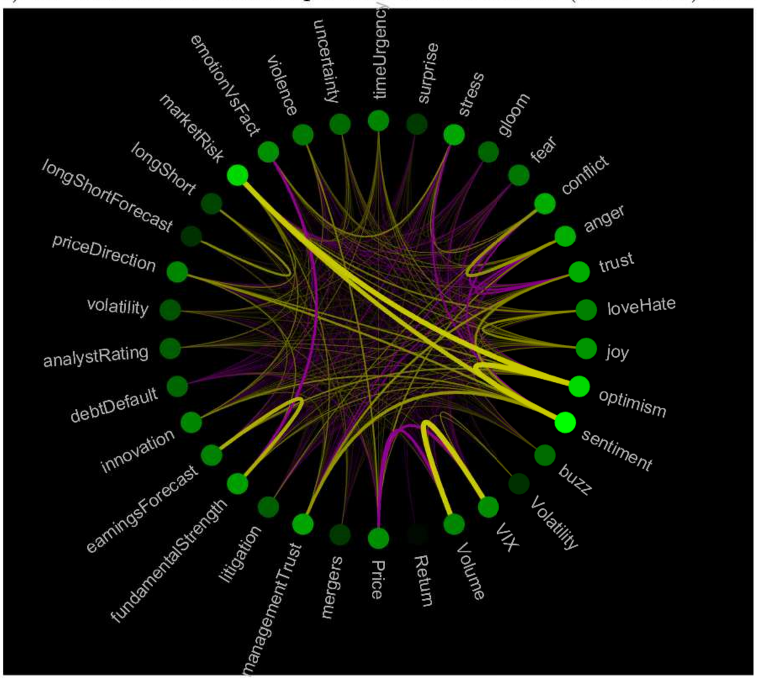


Figure 2: Schema Ball Correlations of TRMI S&P 500 Company Groups. The two panels are visual representation of the pairwise contemporaneous correlations between all 35 scores in the S&P 500 company group, with social based and news media based scores in panel (a) and panel (b) respectively. TRMI indices are complemented by stock market variables, namely, S&P 500 market Prices, Return, Volatility (one month realized volatility), VIX, and Volume. Sample period: 01/Jan/2011 to 30/Nov/2017 at daily frequency. Yellow curves show positive correlation coefficients, purple curves indicate negative correlations, the thickness and brightness of curves represent strength of correlations.

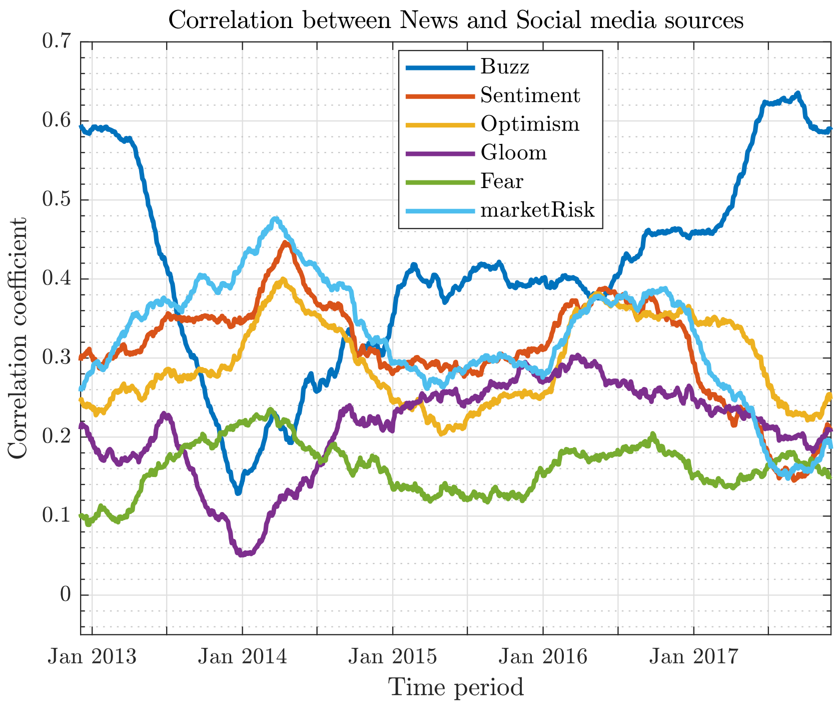


Figure 3: CONTEMPORANEOUS CORRELATION BETWEEN KEY S&P 500 SOCIAL AND NEWS INDICES. All six sentiment indices are from S&P 500 company group from 01/Jan/2011 to 30/Nov/2017, Kendall correlation coefficients are calculated using rolling 500-day estimation window.

correlation between yesterday's news indices with today's social indices. The shape along time horizon are quite similar to contemporaneous correlations above: positively correlated with each other between social and news for all six indices, but the magnitude of correlation coefficients are lower with one day lag and slightly different between panel (a) and panel (b) after July 2014. These two graphs indicate that the causal relationships between social and news media indices may have changed along our sample period, and suggest further causal relationship investigations.

3.5 Principal Component Analysis

Given 35 textual sentiment indices for each TRMI news and social groups, we want to know which measure(s) is/are most representative. The resulting dimensionanlity reduction is imperative in our analysis of VAR and SVAR systems. First, we separate these indices into two groups: polarized group [-1,1] and unidirectional group [0,1] as shown in Table 1 and Table 2 panel (a) and (b). Secondly, we run Principal Component Analysis (PCA) to abstract the top two principal components for polarized and unidirectional sentiment scores respectively. Results of PCA are presented in Figure 4. Note that because *Buzz* metrics are at conceptually different scale to other emotional scores, we do not incorporate *Buzz* in our analysis. Panels (a) and (b) of Figure 4 depict PCA for S&P 500 Index social sentiment indices of [-1,1] and [0,1] groups, respectively, whilst panels (c) and (d) are news sentiment PCA results. Panel (a) and (c) indicate that *Sentiment* and $emotionVsFact$ are most prominent two components in both social and news polarized groups. Comparing panel (c) with panel (d) we find more heterogeneity for unidirectional sentiment measures between social and news indices. Social media emotions tend to be more diverse with *violence*, *stress*, *joy*, *anger* and *gloom* all contributing substantially to overall variability of the dataset. By contrast, *violence* is the dominant emotion in news media. Similar PCA results for S&P 500 social and news indices without differentiating polarized and unidirectional groups are provided in Figure A.2 in the appendix. We find that mixing all polarized and unidirectional scores, *Sentiement* and $emotionVsFact$ are still the most prominent factors.

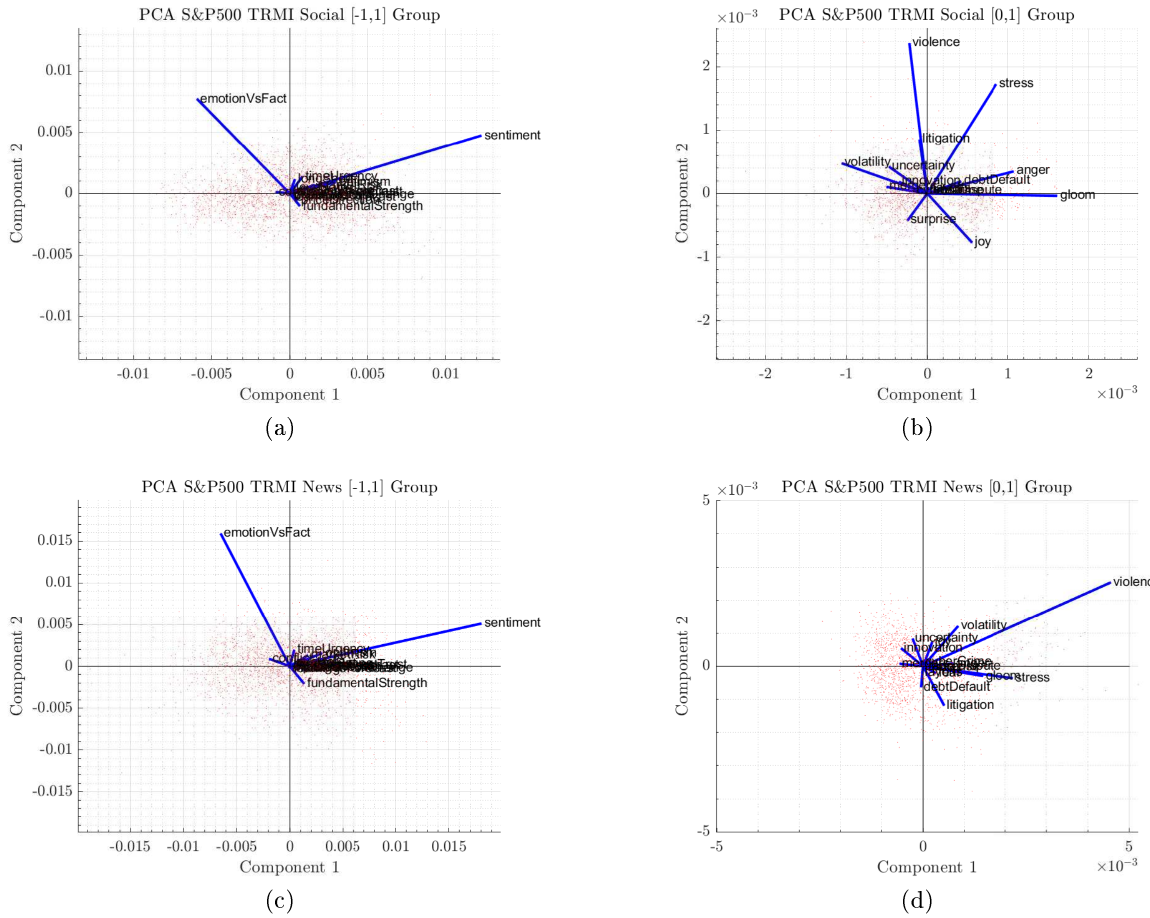
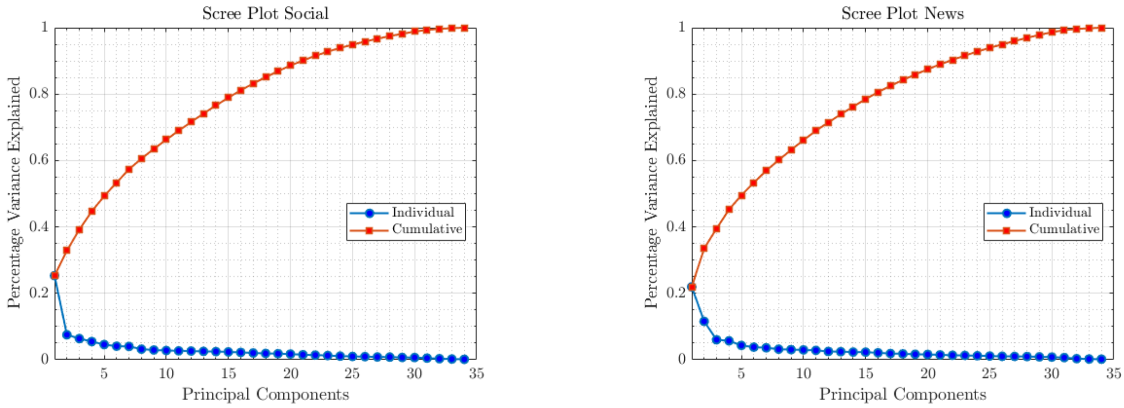


Figure 4: Principal Component Analysis of S&P 500 Sentiment Indices. Panel (a) is a biplot of the first two principal components for the [-1,1] sentiment score group in S&P 500 social sentiment indices; Panel (b) is a biplot of the first two principal components for the [0,1] sentiment score group in S&P 500 social sentiment indices. Panel (c) and (d) are biplots constructed in a similar manner but using news sentiment data instead of social media.

To figure out how many principal components should be considered, we generate scree plots for social and news groups respectively in Figure 5. The left panel shows that the first component explains 25.33% of total social group variance, and the second primary component explains an additional 7.5%. The "kink" happens at the second component, indicating that after the second principal component, incremental explanatory power of other components is greatly diminished. The right panel presents that the first component explains 21.94% of total news group variance, and the second primary component explains additional 11.56%. The "elbow" point appears at the third component for news groups.



(a) Social scree plot

(b) News scree plot

Figure 5: SCREE PLOT OF MPTRXUS500 SENTIMENT INDICES. Panel (a) shows individual (blue curve) as well as cumulative (red curve) contributions of each of the 34 components considered based on PCA of all 34 social sentiment indices; the first component explains 25.33% of total variance, and the second component explains an additional 7.5%. Panel (b) is constructed in a similar manner but based on news sentiment indices. Here, the first component explains 21.94% total variance, and the second component explains additional 11.56%.

3.6 Econometric Framework

To capture linear interdependence we use vector autoregressive (VAR) model as our main research method.[10] VAR provides a simple framework systematically capturing rich dynamics in multiple time-series. We rely on two derivative frameworks: a rolling-window VAR method and structural VAR (SVAR) model to investigate our main research questions, respectively:

1. How social and news media interact with each other overtime?
2. What are the dynamic relationships between media activities and stock market activities?

Generally, to identify a group of simultaneous equation models, one has to make assumptions about endogeneity of the variables considered: which variables are deemed endogenous while others are purely exogenous? These decisions are often criticized as being too subjective (Gujarati, 2009). VAR overcome this shortcoming since it does not assign any prior distinction between endogenous and exogenous variables, i.e. all variables in VAR are endogenous. Thus, we adopt a VAR framework to answer the first research question stated above. A full list of variables used in this study and their definition is available in Table A.3 of the appendix.

Example 1 *In financial time-series econometrics, a multivariate time series* $\mathbf{x}_t$ *is a VAR process of order 1, or VAR(1) for short, if it follows the model:*

$$\mathbf{x}_t = \phi_0 + \mathbf{\Phi} \cdot \mathbf{x}_{t-1} + \epsilon_t$$

where ϕ_0 *is a k-dimensional vector,* $\mathbf{\Phi}$ *is a* $k \times k$ *matrix, and* $\{\epsilon_t\}$ *is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix* Ω.[11] *For instance,* $\mathbf{x}_t$ *could consist of any number of the following variables:*

- *market features (e.g., return, volume, and/or volatility);*
- *TRMI social indices (e.g., buzz, sentiment and/or fear);*
- *TRMI news indices (e.g., buzz, sentiment, gloom, etc.);*

[10] Sims (1980) advocated VAR models as providing a theory-free method to estimate linear interdependence among time-series and to avoid the "incredible identification restrictions".

[11] $\{\epsilon_t\}$ is also called impulse, or innovations (Tsay, 2005).

$\mathbf{x}_t$ *can be generalized to VAR(p), where p is the number of lags considered. To choose the appropriate lag length, p, we use the Akaike Information Criterion(AIC) as well as its variants (e.g. Bayesian information criterion, BIC / Schwartz Criterion, BSC), instead of comparing the* i*-th and* $(i-1)$*-th order VAR model using* $M(i)$ *statistics.*[12]

4 Empirical Results

In this section we focus our attention on causality relationship between news and social media indices and their impacts on market variables.

4.1 News vs Social Media: Causality Analysis

As mentioned in the previous section, to avoid any prior assumptions on variables' exogeneity, we adopt VAR framework in our analysis of news and social dependencies. We examine the serial dependence between $Buzz_N$ and $Buzz_S$ by estimating a VAR specified in systematic equations (1) using DJIA TRMI company group data. We choose DJIA because it is comprised of the largest size and most famous listed companies. As a result, media coverage for this group of companies will suffer less from noise contamination and data sparsity compared with other company groups. Based on our baseline VAR model in Example 1, we represent $k = 2, \mathbf{x}_t = (Buzz_N, Buzz_S)'$, and rewrite the model in scalar form to help with interpretation of coefficients in our subsequent discussion.

$$\begin{aligned} Buzz_{N,t} &= \phi_{N,0} + \Phi_{1,1} Buzz_{N,t-1} + \Phi_{1,2} Buzz_{S,t-1} + \epsilon_{1,t} \\ Buzz_{S,t} &= \phi_{S,0} + \Phi_{2,1} Buzz_{N,t-1} + \Phi_{2,2} Buzz_{S,t-1} + \epsilon_{2,t} \end{aligned} \tag{1}$$

Here, $\Phi_{1,2}$ denotes the linear dependence of $Buzz_{N,t}$ on $Buzz_{S,t-1}$ with lagged dependent variable $Buzz_{N,t-1}$ also as a regressor, so $\Phi_{1,2}$ captures the conditional effect of $Buzz_{S,t-1}$ to $Buzz_{N,t}$ given $Buzz_{N,t-1}$. Analogous interpretation for $\Phi_{2,1}$ also applies by substituting news and social respectively.

Gujarati (2009) distinguishes four cases for such VAR system:

1. Unidirectional causality from $Buzz_S$ to $Buzz_N$ if $\Phi_{1,2}$ is significantly different from zero while $\Phi_{2,1}$ is **NOT** significantly different from zero;
2. Inverse unidirectional causality from $Buzz_N$ to $Buzz_S$ if $\Phi_{2,1}$ is significantly different from zero while $\Phi_{1,2}$ is **NOT** significantly different from zero;
3. Feedback, or bilateral causality, when **both** $\Phi_{1,2}$ and $\Phi_{2,1}$ are significantly different from zero;
4. Independence, when **neither** $\Phi_{1,2}$ nor $\Phi_{2,1}$ are significantly different from zero.

We are more interested in the off-diagonal regression coefficients than the diagonal element coefficients because the level and significance of VAR off-diagonal coefficients tell us different causal relationships, while diagonal elements only show autocorrelation effects.

To perform a rolling-window analysis, we use the past 730 days (a two-year period) as an estima-

[12] For notation and definition details, refer to Table A.3 in the appendix.

tion window. We obtain off-diagonal elements of slope coefficients (Φ_{12} and Φ_{21}) and test their significance. We continue in a similar manner, repeating our analysis each day for the remainder of the sample (898 days). Figure 6 presents the results of this procedure. Each vertical pair of observations represents off-diagonal VAR(1) model slope coefficients. Statistically significant results are emphasised with dots.[13]

From Figure 6, we observe that the blue and red coefficients crossed each other at February 2014. Before this point, the magnitude of red line (Φ_{21}) is above blue line (Φ_{12}), with most of Φ_{21} coefficients being significant and all of the Φ_{12} coefficients insignificant. This phenomenon suggests news media activity is statistically significantly dominating social media activities before February 2014. After this "flip-point", we observe that the absolute values of blue coefficients exceed the red coefficients. During 2014, there are periods that both blue and red coefficients are significant, indicating news and social media activity mutually Granger cause each other. We interpret this period as a coincidence with the SEC's permission to new format media announcements as mentioned in Section 1. Lastly, we find that after January 2015, Φ_{12} (blue, social to news) is trending further upward and remaining significant, while Φ_{21} (red, news to social) is trending downward. This result indicate that social media became dominating news media after January 2015. Overall, this plot shows that the information landscape and market conditions have changed as news media's effects decreasing and social media effects increasing.

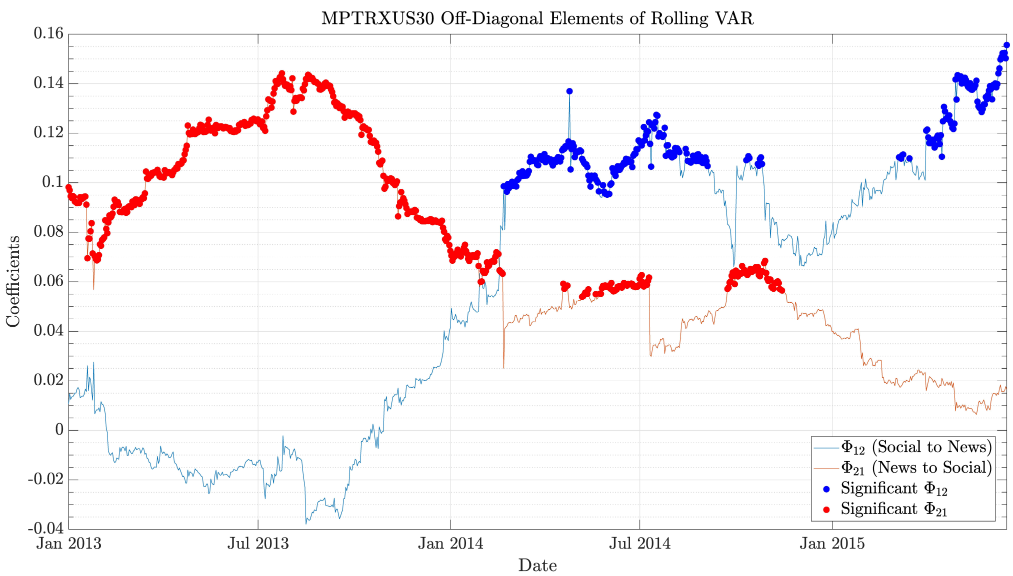


Figure 6: Rolling Window VAR(1) Off-Diagonal Elements of DJIA. This plot depicts interrelationships between company group MPTRXUS30 daily (1440 minutes/ 24 hours) $Buzz_{Social}$ and $Buzz_{News}$ series from 01/Jan/2011 to 30/Nov/2017. The two series have not been merged with market variables. Our sample contains 2,526 observations for each series, and we use the first 730 observations as pre-estimation window, use the last 898 days as estimation window, thus, the rolling window length is also 898 days. Red line represent lead effect from news media to social media, and blue line indicates lead effect from social to news. Significance level above 90% coefficients are marked in dots.

[13] Based on our analysis, a VAR model with 5 lags is optimal according to AIC criterion. However, we report VAR(1) for simplicity of interpretation. Similar rolling window VAR(1) approach was used in DeMiguel et al. (2014) in investigating interrelationship between size portfolios. The results of our VAR(5) model are available upon request.

4.2 Full Sample Structural VAR Results

One of the problems of conventional VAR models is that it does not utilise prior theoretical ideas about how these variables are expected to be related (the "a-theoretical" model specification problem by Hamilton (1994)). Another common criticism comes from the lost of degree of freedom in estimating long lagged multivariate time-series when using unrestricted VAR models. To deal with these two disadvantages, econometricians propose to use structural VAR (SVAR) models that restrict endogenous and exogenous variables based on certain underlying assumptions. A SVAR uses economic theory to sort out the contemporaneous linkages among the variables (Sims et al., 1986; Blanchard and Watson, 1986; Stock and Watson, 2001).

Previous empirical research (Brown and Cliff, 2004; Tetlock, 2007; Antweiler and Frank, 2004) indicates that volume of media activity and sentiment (both positive and negative) are two influential factors to market return, volume and volatility. Assuming both news and social media information have effect on stock market, we test the hypothesis that media activity volume in news and social platforms, proxied by $Buzz_N$ and $Buzz_S$, have different effects on market variables. We further test the hypothesis that net sentiment expressed in news and social media, as measured by $Sent_N$ and $Sent_S$, display different associations with market variables. The reason that we choose *Sentiment* among all 34 TRMI indices lies on the results of previous Principal Component Analysis. Therefore, we restrict our general VAR model in the form: $k = 3, \mathbf{x}_t = (MV, Buzz_N, Buzz_S)'$, where MV represent **one of** our interested market variables: return (r_t), volume (vo_t), realized volatility (σ_t) and VIX (V_t), i.e., $MV = (r_t, vo_t, \sigma_t, V_t)'$. We add VIX into consideration because we believe VIX represents leading volatility and need to be differentiated from realized volatility, which is a trailing volatility.

To deal with the absolute scale differences amongst our raw sample time-series in SVAR, and to avoid the subsequent differences of innovations/shocks from different scales of error terms, we standardize all time-series to have mean zero and variance equal to one, so that innovations/shocks from the dynamic system is not contaminated by scale differences.[14]

For example, a *Return*-$Buzz_N$-$Buzz_S$ SVAR(1) is just a reduced form dynamic structural model specified as:

$$\begin{aligned} r_t &= \phi_{1,0} + \Phi_{1,1} r_{t-1} + \Phi_{1,2} Buzz_{N,t-1} + \Phi_{1,3} Buzz_{S,t-1} + \epsilon_{1,t} \\ Buzz_{N,t} &= \phi_{2,0} + \Phi_{2,1} r_{t-1} + \Phi_{2,2} Buzz_{N,t-1} + \Phi_{2,3} Buzz_{N,t-1} + \epsilon_{2,t} \\ Buzz_{S,t} &= \phi_{3,0} + \Phi_{3,1} r_{t-1} + \Phi_{3,2} Buzz_{N,t-1} + \Phi_{3,3} Buzz_{S,t-1} + \epsilon_{3,t} \end{aligned} \tag{2}$$

Similarly, a *Return*-$Sent_N$-$Sent_S$ SVAR(1) is a reduced form dynamic structural model specified as:

$$\begin{aligned} r_t &= \phi_{1,0} + \Phi_{1,1} r_{t-1} + \Phi_{1,2} Sent_{N,t-1} + \Phi_{1,3} Sent_{S,t-1} + \epsilon_{1,t} \\ Sent_{N,t} &= \phi_{2,0} + \Phi_{2,1} r_{t-1} + \Phi_{2,2} Sent_{N,t-1} + \Phi_{2,3} Sent_{S,t-1} + \epsilon_{2,t} \\ Sent_{S,t} &= \phi_{3,0} + \Phi_{3,1} r_{t-1} + \Phi_{3,2} Sent_{N,t-1} + \Phi_{3,3} Sent_{S,t-1} + \epsilon_{3,t} \end{aligned} \tag{3}$$

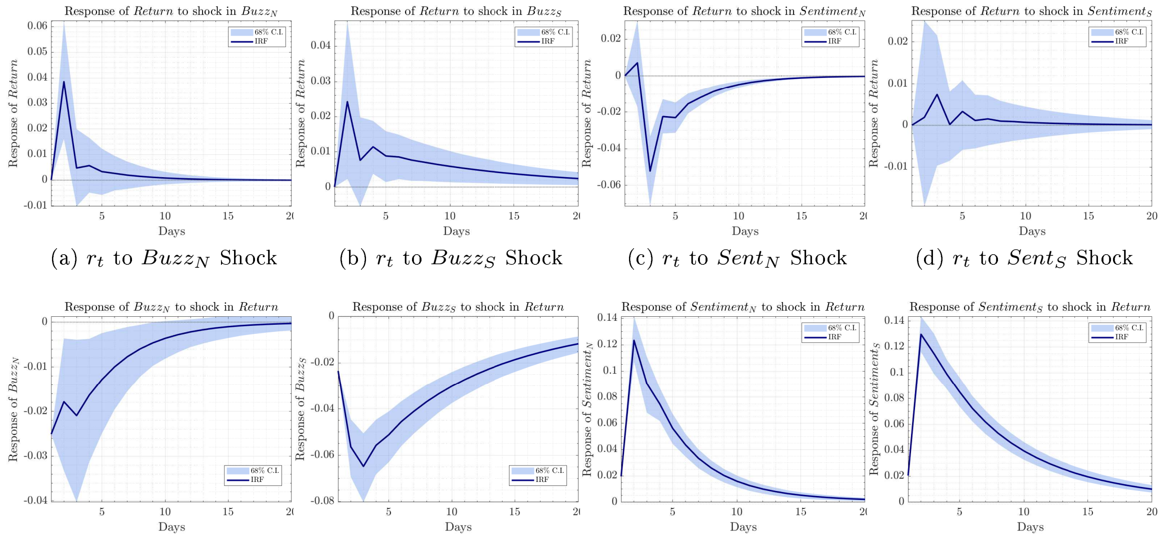
Figure 7 reports the Impulse Response Functions (IRFs thereafter) for our SVAR as demonstrated in systematic equations (2) and (3) above. We continue this dynamic structural model

[14] See examples from Lutz (2015) and Chiu et al. (2018).

estimations by representing in vo_t(Figure 8), V_t (Figure 9), and σ_t (Figure 10) into where r_t is to probe the bilateral causal relationships between *Buzz*, *Sentiment* and market features.

In Figure 7, panel (a) and (b) show return responses to one standard deviation shocks of $Buzz_N$ and $Buzz_S$, respectively. Panel (c) and (d) are return responses to one standard deviation innovations from $Sent_N$ and $Sent_S$ respectively. Panel (e) to (h) represent $Buzz_N$, $Buzz_S$, $Sent_N$ and $Sent_S$ reactions to one standard deviation of unexpected increase in return, i.e., the second row of Figure 7 panels are feedback effects from market to media activities. Blue shaded area are 68% confidence bands for the IRFs following Sims and Zha (1999). From the first row of Figure 7, we find that market return is more sensitive to news media coverage and sentiment than to social media. For example, an unexpected one-unit increase in news media volumes significantly causes market index return to spike 3.8% in the next two days, whilst a one-unit positive surprise in social media activities only significantly causes market return to rise 2.4% in the next two days. An unexpected one-unit positive shock in news media sentiment significantly causes return to decrease 4.5% in the next three days, but return responses to social media sentiment shocks are not statistically significant. The second row of Figure 7 (panel (e) to (h)) reveals the fact that feedback effect from return to news and social media are similar. Rises in return uplift both social and news media sentiment. An one standard deviation surge in return shock significantly increases news media sentiment by 12% in the next two days (panel (g)), and an unexpected one-unit return increase causes social media sentiment to increase by 13% significantly in the next two days (panel (h)). But increases in return lead to less coverage in both news and social media. An one standard deviation positive return shock significanly generates 2.5% less news media reports in the next day (panel (e)), and an one-unit positive return innovation significantly decreases social media discussion volumes by 6.2% in three days (panel (f)). These findngs are consistent with Sprenger et al. (2014) and Brown and Cliff (2004), who also find that social media sentiment do not predict return, but return help predict future social media sentiment.

We continue this SVAR analysis approach to other index characteristics: volume and volatility. Figure 8 report dynamic relationships between volume and news/social buzz and sentiment. Similar to the return SVAR, sensitivities of volume towards news and social media impacts also display heterogeneity. For example, an one-unit positive unexpected news buzz impacts volume negatively in three days, and this effect persist for about one month (20 working days, see panel (a) of Figure 8). In contrast, an one-unit positive unexpected social buzz impacts volume positively in three days and the persistent period also lasts for about one month (panel (b)). Similar discrepencies in volume responses toward news and social media sentiment shocks are also found. In panel (c), market index volume react positively toward unexpected news media sentiment in about four days and then this effect evaporates, while in panel (d) trading volume responds negatively in one day toward one-unit unexpected increase in social media sentiment, in other words, panel (d) shows that unexpected negative sentiment in social media leads to higher amount of trading volume in stock market. This finding is similar to Tetlock (2007) that increase in negative sentiment causes trading volume. In an unreported quintile sort result, we find that $Sent_N$ climbs up with $Buzz_N$ quintiles, while $Sent_S$ is drops down with $Buzz_S$ quintiles, a fact suggesting that news media are biased at reporting good news, but investors talk more nega-



(a) r_t to $Buzz_N$ Shock (b) r_t to $Buzz_S$ Shock (c) r_t to $Sent_N$ Shock (d) r_t to $Sent_S$ Shock

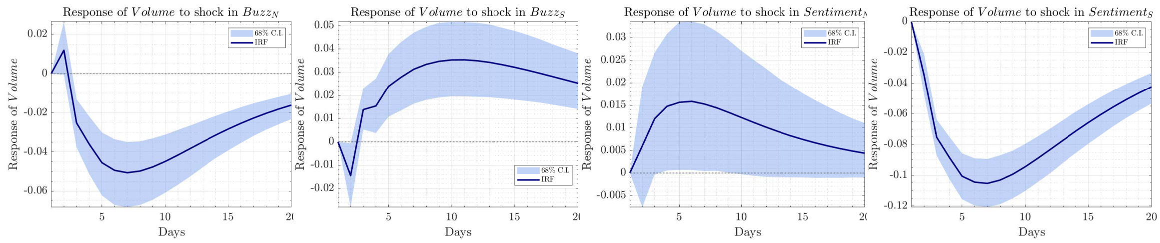
(e) $Buzz_N$ to r_t Shock (f) $Buzz_S$ to r_t Shock (g) $Sent_N$ to r_t Shock (h) $Sent_S$ to r_t Shock

Figure 7: IRFs for News vs Social: Return. This figure contains impulse response functions for r_t-$Buzz_N$-$Buzz_S$ SVAR and r_t-$Sent_N$-$Sent_S$ SVAR. Panel(a) and (b) are return responses for $Buzz_N$ and $Buzz_S$ shocks, respectivle; panel (e) and (f) are feedback effects from return shocks to $Buzz_N$ and $Buzz_S$. Similarly, Panel(c) and (d) are return responses for $Sent_N$ and $Sent_S$ shocks, respectivle; panel (g) and (h) are feedback effects from return shocks to $Sent_N$ and $Sent_S$. All time-series within SVAR are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Sample period: 01/Jan2011-30/Nov/2017.

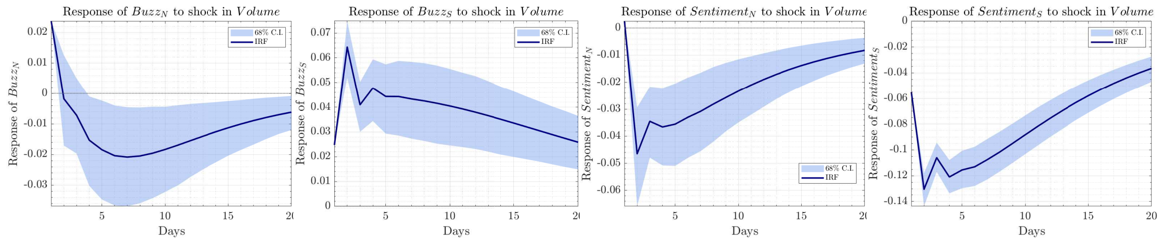
tively in social media platforms. This quintile pattern helps to explain the differences in volume sensitivity to news and social media sentiment. The second row in Figure 8 also provide evidence that feedback effects in news and social media toward volume shocks are consistent with each other. Both news and social media buzz response positively toward one-unit unexpected volume up, although news buzz reaction disappears in approximately two days time with social buzz reaction persist positive. Panel (g) and (h) present that unexpected volume surge causes both news and social media sentiment to react negatively.

Figure 9 and Figure 10 report the SVAR for dynamic relationships between market volatility and news/social activities. We treat VIX (V_t) as forward-looking volatility and realized volatility (σ_t) as backward-looking volatility. Row one of Figure 9 shows that VIX react differently toward news and social buzz and sentiment. Panel (a) reveals that unexpected increase in news media coverage suppress VIX in two days and this decreased level for VIX persists. In contrast, panel (b) shows that reaction of VIX to social media discussion volumes initially decreases but revert back positively, however, this effect is insignificant. Comparing panel (c) and (d), we find that VIX sensitivity to news sentiment shocks is positive and significant in three-day lags and it persists, whilst VIX sensitivity to social sentiment shocks is negative and significant at two-day delays , which also lasts. Panel (g) and (h) of Figure 9 provide further evidence that feedback effects from news and social media sentiment to VIX are consistent.

Different from VIX SVAR, we find that market realized volatility is sensitive to both news and social buzz. To be specific, increases in news and social media coverage shocks significantly decrease volatility in three to four days period (Panel (a) and (b) in Figure 10). Increases in

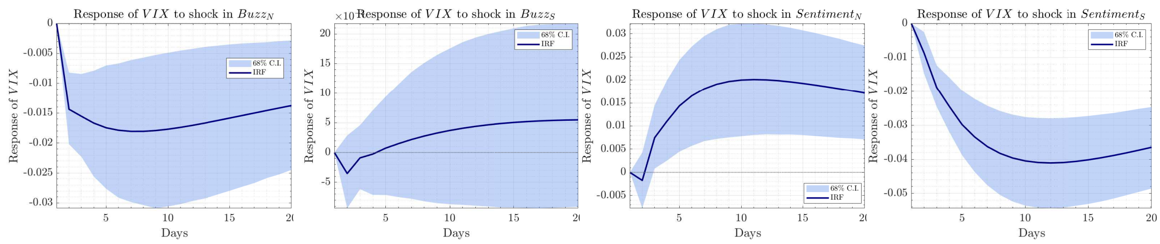


(a) vo_t to $Buzz_N$ Shock (b) vo_t to $Buzz_S$ Shock (c) vo_t to $Sent_N$ Shock (d) vo_t to $Sent_S$ Shock

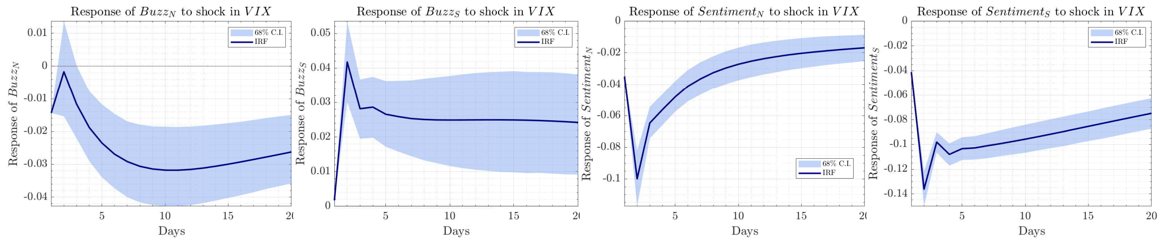


(e) $Buzz_N$ to vo_t Shock (f) $Buzz_S$ to vo_t Shock (g) $Sent_N$ to vo_t Shock (h) $Sent_S$ to vo_t Shock

Figure 8: IRFs FOR NEWS VS SOCIAL: VOLUME. This figure contains impulse response functions for vo_t-$Buzz_N$-$Buzz_S$ SVAR and vo_t-$Sent_N$-$Sent_S$ SVAR. Panel(a) and (b) are volume responses for $Buzz_N$ and $Buzz_S$ shocks, respectivle; panel (e) and (f) are feedback effects from volume shocks to $Buzz_N$ and $Buzz_S$. Similarly, Panel(c) and (d) are volume responses for $Sent_N$ and $Sent_S$ shocks, respectivle; panel (g) and (h) are feedback effects from volume shocks to $Sent_N$ and $Sent_S$. All time-series within SVAR are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Sample period: 01/Jan2011-30/Nov/2017.



(a) V_t to $Buzz_N$ Shock (b) V_t to $Buzz_S$ Shock (c) V_t to $Sent_N$ Shock (d) V_t to $Sent_S$ Shock

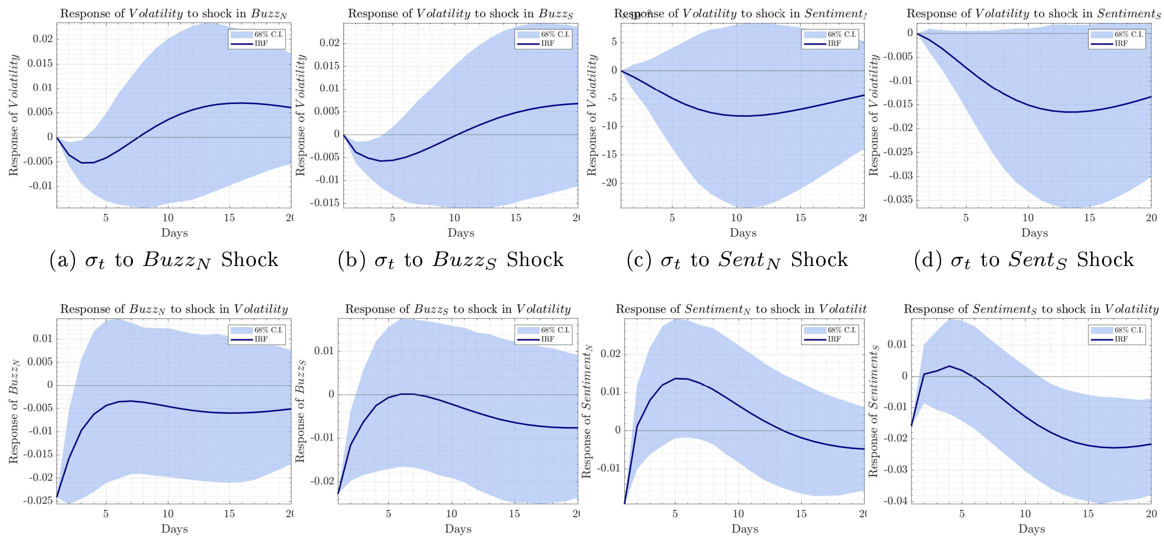


(e) $Buzz_N$ to V_t Shock (f) $Buzz_S$ to V_t Shock (g) $Sent_N$ to V_t Shock (h) $Sent_S$ to V_t Shock

Figure 9: IRFs FOR NEWS VS SOCIAL: VIX. This figure contains impulse response functions for V_t-$Buzz_N$-$Buzz_S$ SVAR and V_t-$Sent_N$-$Sent_S$ SVAR. Panel(a) and (b) are VIX responses for $Buzz_N$ and $Buzz_S$ shocks, respectivle; panel (e) and (f) are feedback effects from VIX shocks to $Buzz_N$ and $Buzz_S$. Similarly, Panel(c) and (d) are VIX responses for $Sent_N$ and $Sent_S$ shocks, respectivle; panel (g) and (h) are feedback effects from VIX shocks to $Sent_N$ and $Sent_S$. All time-series within SVAR are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Sample period: 01/Jan2011-30/Nov/2017.

unexpected news and social positive sentiment also result in decreases in volatility at about two weeks' level. However, sensitivity of volatility to news and social sentiment is insignificant (Panel (c) and (d) in Figure 10). These facts state that market volatility is sensitive to media activeness shocks but insignificant to emotions within these reports. It also implies that information contained in both news and social media help ease the market turmoil at short period (less than a week). Comparing panel (e) with (f), and comparing panel (g) with (h), we obtain supportive evidence that impacts from market variances tend to generate similar effects to both news and social media for both *Buzz* and *Sentiment*. A one-unit increase in volatility shocks creates negative responses in both news and social media coverage, along with significant negative sentiments in both news and social media, at approximately two day's level.

In general, we observe that market responses toward news and social sentiment shocks differ from each other by comparing panel (c) with (d) for each figure from Figure 7 to Figure 10. However, the most prominent finding across all four SVAR figures above appears from panel (g) and (h): that social and news media sentiment responses toward market activity shocks tend to be consistent, and this finding is in line with Jiao et al. (2016a).



(a) σ_t to $Buzz_N$ Shock (b) σ_t to $Buzz_S$ Shock (c) σ_t to $Sent_N$ Shock (d) σ_t to $Sent_S$ Shock

(e) $Buzz_N$ to σ_t Shock (f) $Buzz_S$ to σ_t Shock (g) $Sent_N$ to σ_t Shock (h) $Sent_S$ to σ_t Shock

Figure 10: IRFs for News vs Social: Volatility. This figure contains impulse response functions for σ_t-$Buzz_N$-$Buzz_S$ SVAR and σ_t-$Sent_N$-$Sent_S$ SVAR. Panel(a) and (b) are volatility responses for $Buzz_N$ and $Buzz_S$ shocks, respectivle; panel (e) and (f) are feedback effects from volatility shocks to $Buzz_N$ and $Buzz_S$. Similarly, Panel(c) and (d) are volatility responses for $Sent_N$ and $Sent_S$ shocks, respectivle; panel (g) and (h) are feedback effects from volatility shocks to $Sent_N$ and $Sent_S$. All time-series within SVAR are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Sample period: 01/Jan2011-30/Nov/2017.

4.3 Subsample Structural VAR Results

In section 4.1, we find that news media activeness ($Buzz_N$) has higher influences before February 2014, and social media activeness ($Buzz_S$) tends to have stronger effect after this point of time. In this section, we continue examining how shocks from $Buzz_N$ decrease impacts on market and how shocks from $Buzz_S$ increases impacts on market by applying similar $MV - Buzz_N - Buzz_S$ SVAR analytic approach. We separate our sample into two subsamples based on the "flip-point" found in Section 4.1: subsample 1 from 01/Jan/2013 to 28/Feb/2014 (with 822 observations for each series), subsample 2 from 1/Mar/2014 to 30/Nov/2017 (with 979 observations for each series). We report market variables (r_t, vo_t, V_t, and σ_t) responses to shocks from $Buzz_N$, and $Buzz_S$ for the two subsamples in Figure 11 and Figure 12, respectively. Market sensitivity to $Sent_N$ and $Sent_S$ in two subsamples are offered in Figure A.3 and Figure A.4 in the appendix. Feedback effects at subsample periods are also provided in the appendix.[15]

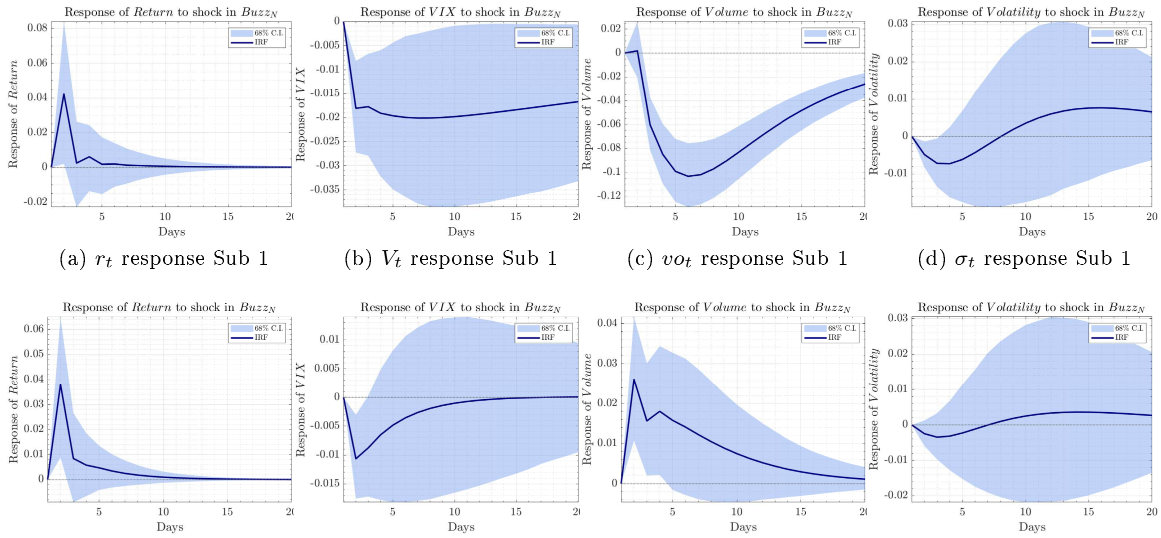
In Figure 11, we find that increases in news quantity shocks increases return at both subsample 1 and subsample 2 and in similar time lags by comparing panel (a) with (e), but the magnitudes in two subsamples are slightly different, with the post-2014 period less sensitive. The diminishing sensitivity to $Buzz_N$ also appears in volatility responses (panel (b) vs (f), and panel (d) vs (h)). However, the persistent negative sensitivity of volatility to $Buzz_N$ shocks changed after the "flip-point": response of trading volume to unexpected increases in news quantities become significantly positive in two to four days' frame. In sum, we observe that return and volatility become less sensitive to $Buzz_N$ shocks, and trading volume sensitivity to $Buzz_N$ flipped-sign after February 2014.

Figure 12 is market responses to $Buzz_S$ shocks at the two different periods. We find that r_t, vo_t and V_t all become more sensitive to social media volume shocks. For example, subsample 1 return response is positive but insignificant (panel (a)). However, it turns to be significant at the second subsample (panel (e)). The insignificant negative response of VIX to $Buzz_S$ shocks (panel (b)) also becomes significant at the second period (panel (f)). In Panel (g), confidence bands of trading volume sensitivity to social media discussions narrows down and moves further away from zero comparing with panel (c), which also suggests that market responses to subsample 2 period become more sensitive. Lastly, an one standard deviation increase in social media postings causes market volatility to decrease in the next two to three days in subsample 1 (panel (d)), but in subsuample 2, impacts from social media shocks make volatility react persistently positive and in a larger magnitude (panel (h)). Overall, this evidence expresses how influences from social media to market variables become stronger after February 2014.

5 Conclusion

In this paper, we use rolling-horizon VAR and structural VAR (SVAR) models to investigate how mere quantities of posts as well as the sentiment identified within these posts impact the US financial markets. We contrast the effects of social vs news media from 2011 to 2017 by applying sentiment scores from Thomson Reuters MarketPsych Indices (TRMI), which extracts texts from major social and news media outlets by a machine-learning algorithm. We find that social media tends to have stronger effects on stock markets than news media does after February 2014 - a period when the SEC admits the thriving social media as official information dissemination channels for public companies. We continue examining and comparing overall market

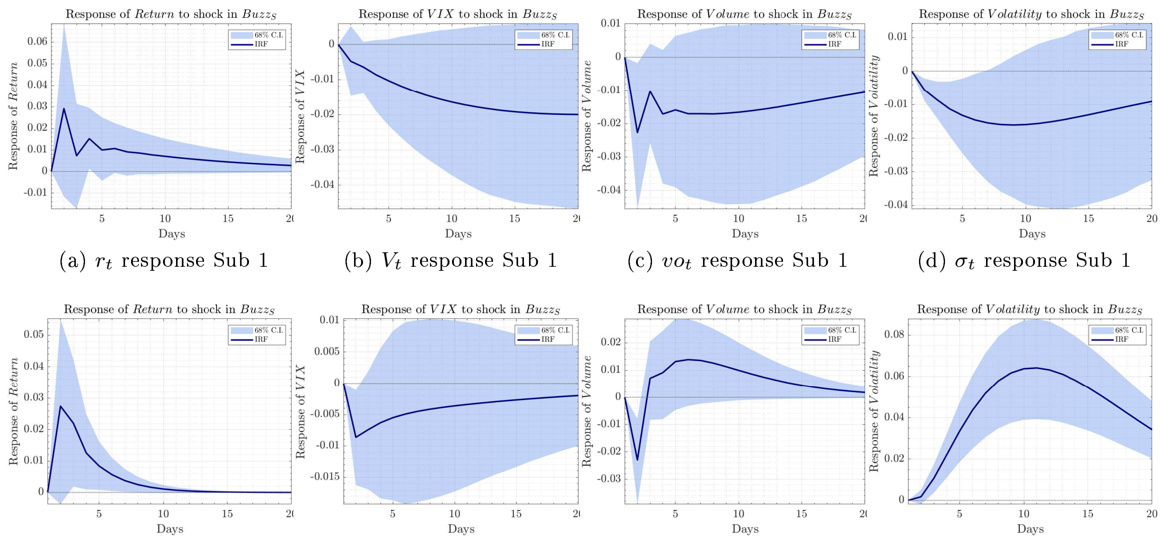
[15] Figure A.5 reports feedback effects from market variable shocks to $Buzz_N$ at the two subsamples, similarly, Figure A.6 provides feedback effects from market variable shocks to $Buzz_S$ in the two subsumples, Figure A.7 is subsample $Sent_N$ reactions, and Figure A.8 is subsample $Sent_S$ reactions.



(a) r_t response Sub 1 (b) V_t response Sub 1 (c) vo_t response Sub 1 (d) σ_t response Sub 1

(e) r_t response Sub 2 (f) V_t response Sub 2 (g) vo_t response Sub 2 (h) σ_t response Sub 2

Figure 11: Market Reactions to $Buzz_N$ Shocks - Subsample Comparison. Panel(a) to (d) depict IRFs of market variable reactions to one standard deviation shock in $Buzz_N$ in subsample 1, panel(e) to (h) are generated in the same manner but for subsample 2. All time-series variables are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.



(a) r_t response Sub 1 (b) V_t response Sub 1 (c) vo_t response Sub 1 (d) σ_t response Sub 1

(e) r_t response Sub 1 (f) V_t response Sub 2 (g) vo_t response Sub 2 (h) σ_t response Sub 2

Figure 12: Market Reactions to $Buzz_S$ Shocks - Subsample Comparison. Panel(a) to (d) depict IRFs of market variable reactions to one standard deviation shock in $Buzz_S$ in subsample 1, panel(e) to (h) are generated in the same manner but for subsample 2. All time-series variables are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.

sensitivities (sensitivity of return, volume and volatility) to shocks from news and social media quantities (*Buzz*) and sentiment (*Sent*). Consistent with past empirical literature that use textual analysis data to test market predictability, we find significant heterogeneity in market reactions to news and social media shocks. In contrast, news and social media reactions to (feedback effects from) unexpected changes in return, volume and volatility are homogeneous and consistent. We also find that returns could forecast social media sentiment but social media sentiment shows insignificant predictability in our full sample period. We find that news media sentiment and S&P 500 short-period returns mutually influence each other, and that negative news media sentiment causes S&P 500 turnover. Subsample SVAR analysis on pre and post February 2014 - the "flip-point" that we find in our rolling-window VAR analysis corroborates and further explains how news media has less effects while social media has a stronger impact on market variables. Overall, this paper reveals new insights about effects of social and news media on markets, identifies dynamic relationships they have with the market, and helps shed light on how quantities and content of information in news and social media influence stock markets.

Our findings also bring up some interesting future research questions. We aim at exploring the question of why differences exist in the market sensitivity to news and social media shocks. We have also noticed that a lot of trading activities relying on textual analysis sentiment are closed out within the trading day. This brings forth the question of how news and social media sentiment impact stock market on a high-frequency (intraday) basis. We believe investigations using this novel data will help us contribute to literature on return predictability from sentiment and literature on how sentiment impact on stock performances.

References

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294.

Audrino, F. and Tetereva, A. (2017). Sentiment spillover effects for us and european companies.

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2):129–152.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment1. *Journal of Financial Economics*, 49(3):307–343.

Bartov, E., Faurel, L., and Mohanram, P. S. (2018). Can twitter help predict firm-level earnings and stock returns? *The Accounting Review*, 93(3):25–57.

Black, F. (1986). Noise. *The Journal of Finance*, 41(3):528–543.

Blanchard, O. J. and Watson, M. W. (1986). Are business cycles all alike? In *The American Business Cycle: Continuity and Change*, pages 123–180. University of Chicago Press.

Brown, G. W. and Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1):1–27.

Campbell, J. Y. and Kyle, A. S. (1993). Smart money, noise trading and stock price behaviour. *The Review of Economic Studies*, 60(1):1–34.

Chen, H., De, P., Hu, Y. J., and Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5):1367–1403.

Chiu, C.-W. J., Harris, R. D., Stoja, E., and Chin, M. (2018). Financial market volatility, macroeconomic fundamentals and investor sentiment. *Journal of Banking & Finance*, 92:130–145.

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

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance*, 53(6):1839–1885.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738.

DeMiguel, V., Nogales, F. J., and Uppal, R. (2014). Stock return serial dependence and out-of-sample portfolio performance. *The Review of Financial Studies*, 27(4):1031–1073.

Engelberg, J. (2008). Costly information processing: Evidence from earnings announcements.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5):2023–2052.

Fehr, E. and Tyran, J.-R. (2005). Individual irrationality and aggregate outcomes. *Journal of Economic Perspectives*, 19(4):43–66.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408.

Gujarati, D. N. (2009). *Basic Econometrics.* Tata McGraw-Hill Education.

Hamilton, J. D. (1994). *Time Series Analysis*, volume 2. Princeton university press Princeton, NJ.

Hardy, C. O. (1939). *Odd-Lot Trading on the New York Stock Exchange.* Brooking Institution, Washington.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973)*, 45(4):363–407.

Heston, S. L. and Sinha, N. R. (2017). News vs sentiment: predicting stock returns from news stories. *Financial Analysts Journal*, 73(3):67–83.

Hirshleifer, D. and Teoh, S. H. (2008). Thought and behavior contagion in capital markets.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6):2143–2184.

Huang, D., Lehkonen, H., Pukthuanthong, K., and Zhou, G. (2018). Sentiment across asset markets.

Jegadeesh, N. and Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3):712–729.

Jiao, P., Veiga, A., and Walther, A. (2016a). Signal processing on social media: Theory and evidence from financial markets.

Jiao, P., Veiga, A., and Walther, A. (2016b). Social media, news media and the stock market.

Kearney, C. and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33:171–185.

Kirkpatrick II, C. D. and Dahlquist, J. A. (2010). *Technical Analysis: the Complete Resource for Financial Market Technicians.* FT press.

Le Bon, G. L. (1896). *The Crowd: A Study of the Popular Mind.* Classic Books Library.

Lee, C. M., Shleifer, A., and Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, 46(1):75–109.

Liu, B. and McConnell, J. J. (2013). The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? *Journal of Financial Economics*, 110(1):1–17.

Loughran, T. and McDonald, B. (2011a). Barron's red flags: Do they actually work? *Journal of Behavioral Finance*, 12(2):90–97.

Loughran, T. and McDonald, B. (2011b). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.

Lutz, C. (2015). The impact of conventional and unconventional monetary policy on investor sentiment. *Journal of Banking & Finance*, 61:89–105.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3):483–510.

Nardo, M., Petracco-Giudici, M., and Naltsidis, M. (2016). Walking down wall street with a tablet: A survey of stock market predictions using the web. *Journal of Economic Surveys*, 30(2):356–369.

Neal, R. and Wheatley, S. M. (1998). Do measures of investor sentiment predict returns? *Journal of Financial and Quantitative Analysis*, 33(4):523–547.

Preis, T., Moat, H. S., and Stanley, H. E. (2013). Quantifying trading behavior in financial markets using google trends. *Scientific Reports*, 3:1684.

Qiu, L. and Welch, I. (2004). Investor sentiment measures. Technical report, National Bureau of Economic Research.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PloS one*, 10(9):e0138441.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the us stock market. *Journal of Banking & Finance*, 84:25–40.

Shen, J., Najand, M., Dong, F., and He, W. (2017). News and social media emotions in the commodity market. *Review of Behavioral Finance*, 9(2):148–168.

Siganos, A., Vagenas-Nanos, E., and Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*, 107:730–743.

Siganos, A., Vagenas-Nanos, E., and Verwijmeren, P. (2017). Divergence of sentiment and stock market trading. *Journal of Banking & Finance*, 78:130–141.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48.

Sims, C. A. et al. (1986). Are forecasting models usable for policy analysis? *Quarterly Review*, (Win):2–16.

Sims, C. A. and Zha, T. (1999). Error bands for impulse responses. *Econometrica*, 67(5):1113–1155.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., and Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5):926–957.

Stock, J. H. and Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.

Swaminathan, B. (1996). Time-varying expected small firm returns and closed-end fund discounts. *The Review of Financial Studies*, 9(3):845–887.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467.

Tsay, R. S. (2005). *Analysis of Financial Time Series*, volume 543. John Wiley & Sons.

Wysocki, P. (1998). Cheap talk on the web: The determinants of postings on stock message boards.

Zweig, M. E. (1973). An investor expectations stock price predictive model using closed-end fund premiums. *The Journal of Finance*, 28(1):67–78.

Appendix

A List of acronyms and notation

Table A.1: List of acronyms.

Acronym	Description
ACF	Autocorrelation Function
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BSC	Schwartz Criterion
CEFD	closed-end fund discount
Datastream	Thomson Reuters Datastream
DJIA	Dow Jones Industry Average
DJNS	Dow Jones Newswires
DW	Durbin-Watson test
GFC	Global Financial Crisis
GI	Harvard General Inquirer Dictionary
GSV	Google Search Volume
IQR	Interquartile Range
IRF	Impulse Response Function
LB	Ljung-Box test
MV	Market Variables
PCA	Principal Component Analysis
RIC	Reuters Identification Code
S&P 100	Standard & Poor's 100 Index
S&P 500	Standard & Poor's 500 Index
SEC	The US Securities and Exchange Commission
SIRCA	Securities Industry Research Centre of Asia-Pacific
SVAR	Structural Vector Autoregressive Model
TR	Thomson Reuters
TRMI	Thomson Reuters MarketPsych Indices
TRNA	Thomson Reuters News Analytics
TRTH	Thomson Reuters Tick History
VAR	Vector Autoregressive Model
WSJ	The Wall Street Journal

B Data sources

Table A.2: List of data sources.

Code	Description
.SPY	SPDR S&P 500 ETF RIC
CBOE	Chicago Board Options Exchange
Datastream	Thomson Reuters Datastream
MPTRXUS30	TRMI company group code (DJIA respective sentiment)
MPTRXUS500	TRMI company group code (S&P 500 respective sentiment)
SIRCA	Securities Industry Research Centre of Asia-Pacific
WRDS	Wharton Research Data Services

C Variable names and definition.

Table A.3: List of variable denote and definition.

Symbol	Description
$Buzz_N$	news media buzz (report volume in news media)
$Buzz_S$	social media buzz (posting volume in social media)
$Sent_N$	news media net sentiment (positive minus negative sentiment)
$Sent_S$	social media net sentiment (positive minus negative sentiment)
r_t	log return on day t
vo_t	trading volume on day t
σ_t	realized volatility over past 22 trading day (1 month) on day t
V_t	VIX (CBOE options volatility index) on day t

D Descriptive Statistics for MPTRXUS30 / DJIA

Table A.4: Descriptive Statistics for TRMI MPTRXUS30 Company Groups Social Indices. Sample period 01/Jan/2011 - 30/Nov/2017; sentiment indices are grouped into polarized scores [-1,1] and unidirectional scores [0,1]. Data in “ - ” were too sparse during our sample period, we include for completeness. Durbin-Watson test and Ljung-Box 5 lags test for all indices show that there is autocorrelation.

	Panel (A): Polarized Groups [-1,1]									
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
sentiment	-0.011	0.035	0.134	-0.173	0.05	3.38	-0.036	-0.012	0.011	0.047
optimism	0.002	0.010	0.041	-0.051	-0.14	3.93	-0.004	0.003	0.009	0.013
loveHate	0.005	0.002	0.022	-0.002	2.44	16.79	0.005	0.005	0.006	0.001
trust	-0.002	0.003	0.004	-0.044	-4.15	43.92	-0.003	-0.002	-0.001	0.003
conflict	0.022	0.006	0.075	0.001	1.22	8.10	0.018	0.021	0.025	0.007
timeUrgency	0.020	0.005	0.058	-0.009	0.89	6.80	0.017	0.020	0.023	0.007
emotionVsFact	0.523	0.028	0.646	0.249	-1.14	13.60	0.508	0.524	0.539	0.032
marketRisk	-0.007	0.005	0.013	-0.036	-0.29	4.16	-0.010	-0.006	-0.003	0.007
longShort	0.004	0.007	0.193	-0.014	19.77	516.20	0.002	0.004	0.006	0.004
longShortForecast	0.001	0.001	0.007	-0.007	-0.46	10.60	0.001	0.001	0.002	0.001
priceDirection	0.002	0.003	0.015	-0.010	0.09	4.54	0.001	0.002	0.004	0.003
priceForecast	0.001	0.001	0.005	-0.003	0.21	6.86	0.000	0.001	0.001	0.001
analystRating	0.001	0.001	0.012	-0.005	1.37	12.29	0.000	0.001	0.002	0.001
dividends	-	-	-	-	-	-	-	-	-	-
earningsForecast	0.002	0.001	0.008	-0.006	0.76	6.96	0.001	0.002	0.002	0.001
fundamentalStrength	0.005	0.003	0.033	-0.010	1.36	9.92	0.004	0.005	0.007	0.003
managementChange	0.002	0.002	0.052	0.000	10.22	191.54	0.001	0.001	0.002	0.001
managementTrust	-0.001	0.005	0.031	-0.103	-10.24	159.55	-0.001	0.000	0.001	0.002
	Panel (B): Unidirectional Groups [0,1]									
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
anger	0.013	0.003	0.074	0.006	5.04	87.07	0.011	0.013	0.015	0.004
fear	0.005	0.001	0.014	0.002	1.46	10.10	0.004	0.005	0.006	0.001
joy	0.014	0.003	0.039	0.006	1.56	10.91	0.012	0.014	0.016	0.004
gloom	0.027	0.005	0.081	0.013	1.01	10.63	0.024	0.027	0.030	0.007
stress	0.053	0.005	0.079	0.034	0.45	5.18	0.050	0.052	0.055	0.006
surprise	0.008	0.002	0.019	0.004	1.20	5.47	0.007	0.008	0.009	0.002
uncertainty	0.024	0.004	0.043	0.011	0.86	4.14	0.021	0.023	0.026	0.004
violence	0.030	0.007	0.088	0.016	2.07	10.02	0.026	0.029	0.033	0.007
volatility	0.026	0.004	0.075	0.016	2.21	21.19	0.024	0.026	0.028	0.005
debtDefault	0.004	0.002	0.034	0.001	4.80	61.35	0.003	0.004	0.005	0.002
innovation	0.003	0.001	0.017	0.001	2.31	18.25	0.002	0.003	0.004	0.001
laborDispute	-	-	-	-	-	-	-	-	-	-
layoffs	-	-	-	-	-	-	-	-	-	-
litigation	0.008	0.003	0.044	0.003	3.15	24.57	0.006	0.007	0.009	0.003
mergers	0.003	0.002	0.019	0.000	3.37	23.22	0.002	0.002	0.003	0.001
cyberCrime	-	-	-	-	-	-	-	-	-	-
	Panel (C): Buzz									
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
buzz	45,149.99	18,353.48	176,650.08	4,239.20	1.56	8.10	32,883.39	42,579.70	52,968.48	20,085.09

Table A.5: Descriptive Statistics for TRMI MPTRXUS30 Company Groups News Indices. Sample period 01/Jan/2011 - 30/Nov/2017; sentiment indices are grouped into polarized scores [-1,1] and unidirectional scores [0,1]. Buzz is a special measure which has different scale from all other metrics. Data in " - " were too sparse during our sample period, we include for completeness. Durbin-Watson test and Ljung-Box 5 lags test for all indices show there is autocorrelation.

				Panel (A): Polarized Groups [-1,1]						
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
sentiment	-0.012	0.041	0.136	-0.169	-0.07	3.25	-0.039	-0.012	0.015	0.054
optimism	0.006	0.009	0.060	-0.036	-0.21	4.47	0.001	0.006	0.011	0.011
loveHate	-	-	-	-	-	-	-	-	-	-
trust	-0.002	0.003	0.010	-0.020	-1.04	7.57	-0.003	-0.002	0.000	0.003
conflict	0.034	0.008	0.082	0.014	0.90	4.80	0.029	0.033	0.038	0.010
timeUrgency	0.025	0.005	0.063	0.006	0.97	6.90	0.022	0.025	0.028	0.006
emotionVsFact	0.523	0.033	0.620	0.386	-0.64	3.69	0.505	0.528	0.546	0.041
marketRisk	-0.007	0.005	0.011	-0.029	-0.40	3.88	-0.010	-0.007	-0.004	0.006
longShort	0.002	0.003	0.026	-0.029	-0.59	12.43	0.000	0.002	0.003	0.003
longShortForecast	0.000	0.001	0.007	-0.004	0.09	10.39	0.000	0.000	0.001	0.001
priceDirection	0.003	0.004	0.022	-0.024	0.06	7.26	0.001	0.003	0.004	0.004
priceForecast	-	-	-	-	-	-	-	-	-	-
analystRating	-	-	-	-	-	-	-	-	-	-
dividends	-	-	-	-	-	-	-	-	-	-
earningsForecast	-	-	-	-	-	-	-	-	-	-
fundamentalStrength	0.008	0.006	0.046	-0.018	1.82	10.31	0.005	0.007	0.009	0.005
managementChange	-	-	-	-	-	-	-	-	-	-
managementTrust	-	-	-	-	-	-	-	-	-	-

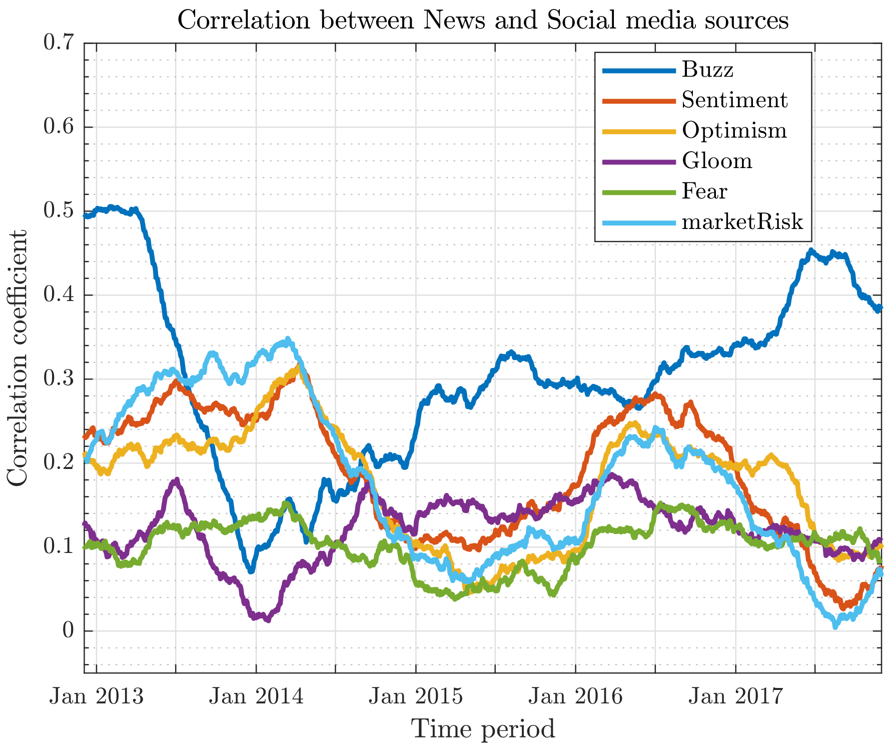
				Panel (B): Unidirectional Groups [0,1]						
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
anger	-	-	-	-	-	-	-	-	-	-
fear	-	-	-	-	-	-	-	-	-	-
joy	-	-	-	-	-	-	-	-	-	-
gloom	0.024	0.004	0.052	0.016	1.63	8.89	0.022	0.023	0.026	0.004
stress	0.055	0.006	0.084	0.035	0.45	4.11	0.052	0.055	0.058	0.007
surprise	0.007	0.002	0.031	0.003	2.42	22.05	0.006	0.007	0.008	0.002
uncertainty	0.020	0.003	0.034	0.010	0.48	4.02	0.018	0.020	0.022	0.004
violence	0.042	0.012	0.123	0.017	2.01	10.18	0.034	0.039	0.046	0.013
volatility	0.031	0.004	0.055	0.020	0.72	5.30	0.029	0.031	0.033	0.004
debtDefault	0.004	0.001	0.014	0.001	1.81	8.80	0.003	0.003	0.004	0.002
innovation	0.006	0.002	0.031	0.000	2.49	28.52	0.004	0.005	0.007	0.002
laborDispute	-	-	-	-	-	-	-	-	-	-
layoffs	-	-	-	-	-	-	-	-	-	-
litigation	0.013	0.005	0.062	0.003	2.02	12.07	0.009	0.011	0.015	0.006
mergers	0.004	0.002	0.021	0.000	2.60	14.88	0.002	0.003	0.004	0.002
cyberCrime	-	-	-	-	-	-	-	-	-	-

				Panel (C): Buzz						
	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
buzz	89,363.69	24,003.80	210,148.99	586.70	0.59	4.76	74,368.71	88,208.40	101,622.70	27,253.99

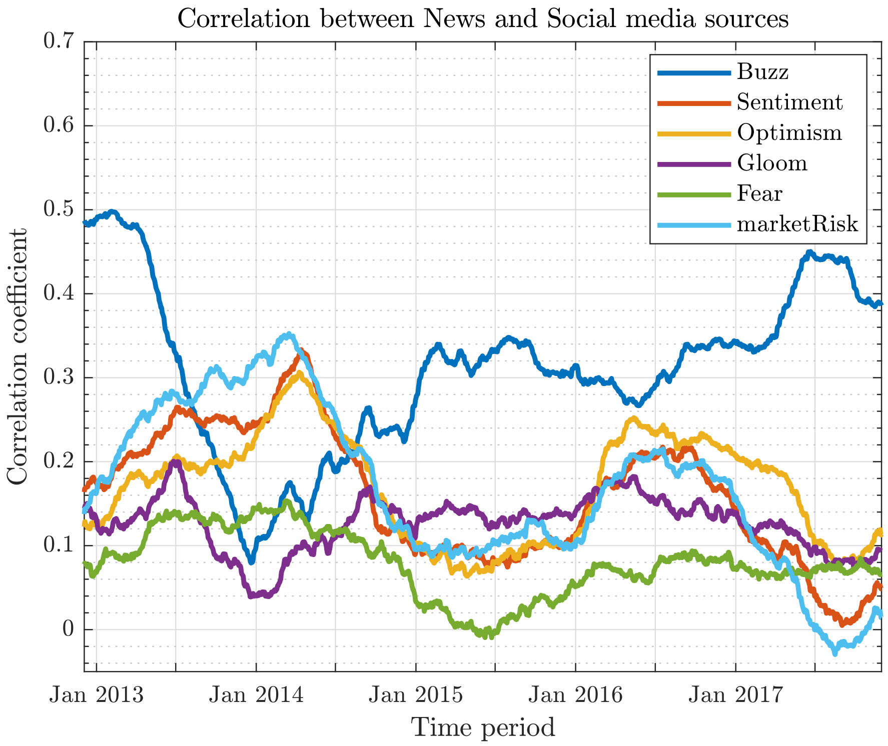
Table A.6: DESCRIPTIVE STATISTICS FOR DOW JONES INDUSTRY AVERAGE. Sample period 01/Jan/2011 - 30/Nov/2017; Return is annualized by multiplying daily values by 252; Volume is scaled at 10^5; Durbin-Watson test and Ljung-Box 5-lag test results indicate that there is autocorrelation.

	Mean	**Std**	**Max**	**Min**	**Skew**	**Kurt**	**25th**	**Median**	**75th**	**ITQ**
Return	0.09	1.87	9.34	-12.8	-0.5	7.68	-0.65	0.07	1.02	1.67
Volume	1.48	0.86	6.03	0.34	1.7	5.84	0.91	1.19	1.67	0.76
VIX	15.37	4.84	41.5	7.58	2.17	8.65	12.4	14.14	16.62	4.21
Volatility	202.9	103.9	683.6	57.4	1.7	6.9	134.0	181.5	244.8	110.7

E One day lag cross correlations between social and news.



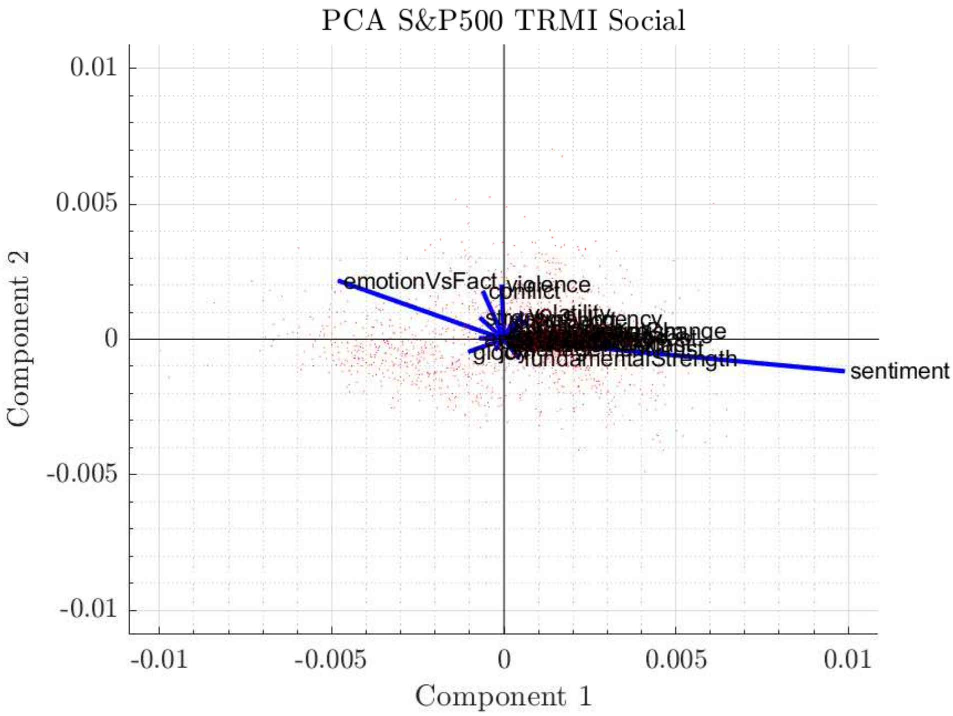
(a) *Social* leads *News* one day



(b) *News* leads *Social* one day

Figure A.1: One Day Lag Cross-Correlation between S&P 500 Key Social and News Indices. Panel (a) shows Kendal correlation between key social and news sentiment indices for TRMI company group MPTRXUS500 daily data, where social leads news one day, i.e. cross-correlation between $Social_t$ and $News_{t-1}$; panel (b) shows Kendal correlation between key news and social sentiment indices data, where news leads social one day, i.e. cross-correlation between $News_t$ and $Social_{t-1}$.

F PCA of 34 TRMI Indices for S&P 500 Company Group



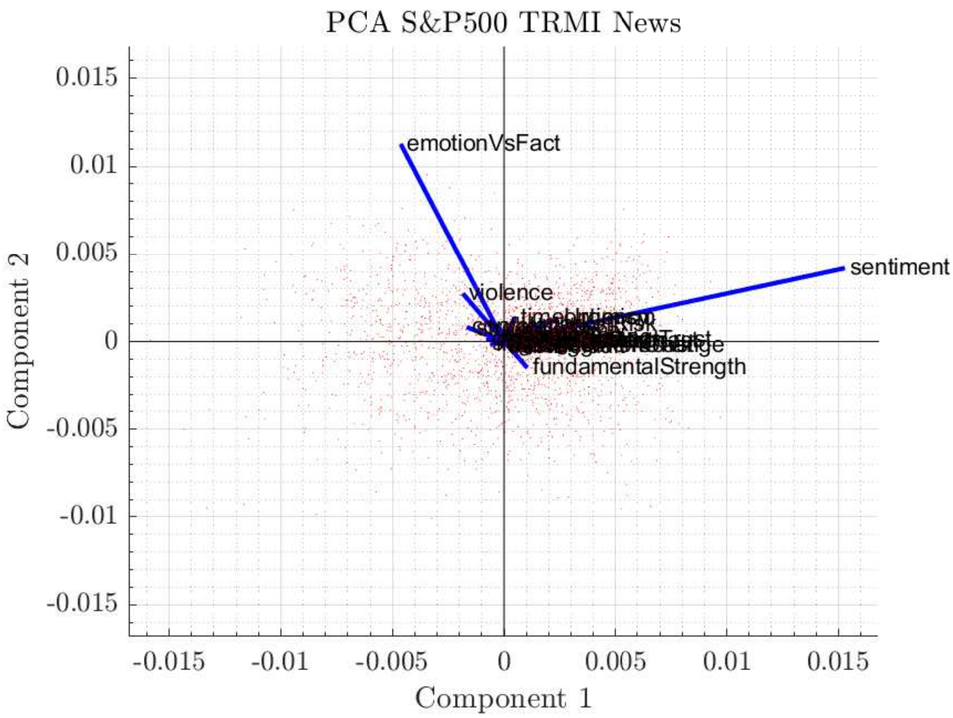
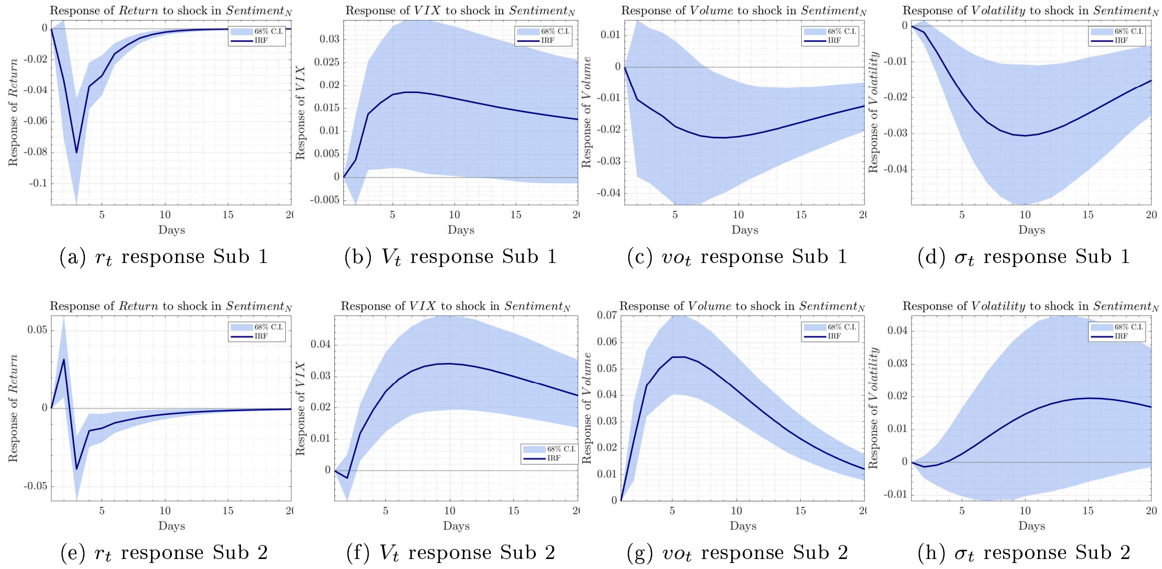


Figure A.2: Principal Component Analysis for S&P 500 TRMI Social and News Indices. Panel (a) is social and Panel (b) is news. $Buzz_{social}$ and $Buzz_{n}ews$ are not included because $Buzz$ metrics are in quite different scale from the other 34 sentiment measures.

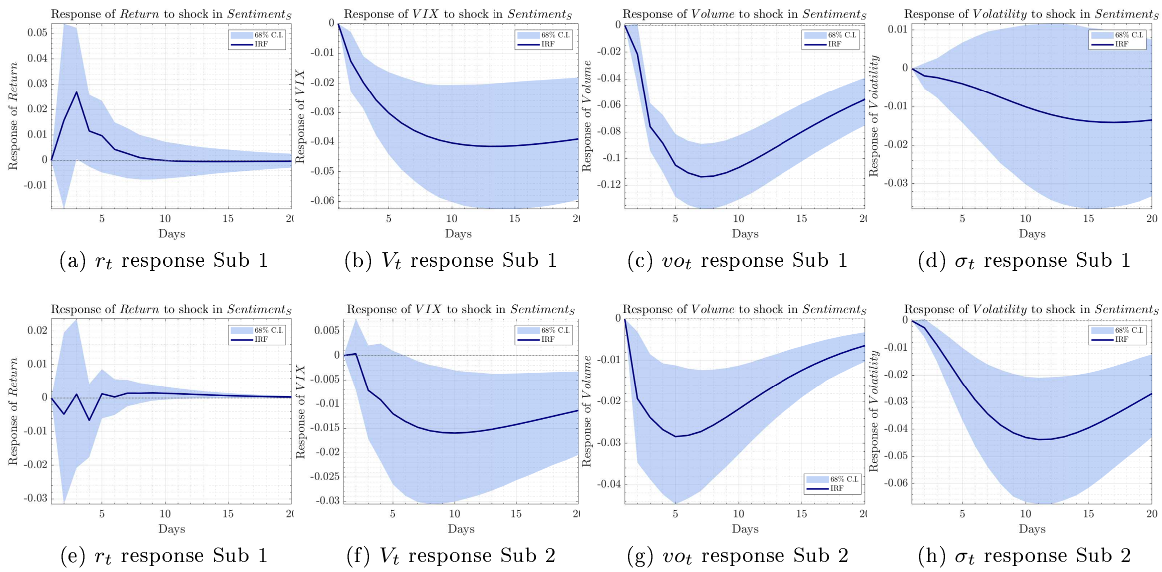
G Subsample Comparison of Sentiment Shocks



(a) r_t response Sub 1 (b) V_t response Sub 1 (c) vo_t response Sub 1 (d) σ_t response Sub 1

(e) r_t response Sub 2 (f) V_t response Sub 2 (g) vo_t response Sub 2 (h) σ_t response Sub 2

Figure A.3: MARKET REACTIONS TO $Sent_N$ SHOCKS - SUBSAMPLE COMPARISON. Panel(a) to (d) depict IRFs of market variable reactions to one standard deviation shock in $Sent_N$ in subsample 1, panel(e) to (h) are generated in the same manner but for subsample 2. All time-series variables are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.

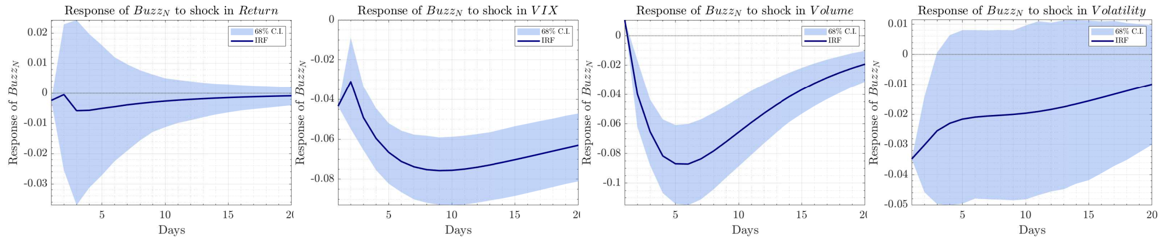


(a) r_t response Sub 1 (b) V_t response Sub 1 (c) vo_t response Sub 1 (d) σ_t response Sub 1

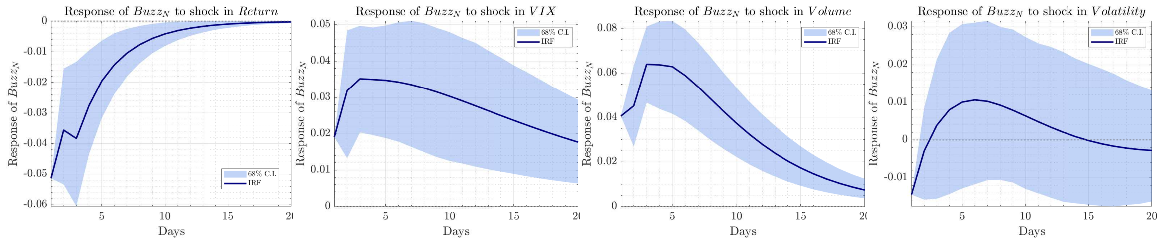
(e) r_t response Sub 1 (f) V_t response Sub 2 (g) vo_t response Sub 2 (h) σ_t response Sub 2

Figure A.4: MARKET REACTIONS TO $Sent_S$ SHOCKS - SUBSAMPLE COMPARISON. Panel(a) to (d) depict IRFs of market variable reactions to one standard deviation shock in $Sent_S$ in subsample 1, panel(e) to (h) are generated in the same manner but for subsample 2. All time-series variables are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.

H Subsample Comparison of Feedback Effect from Market Shocks

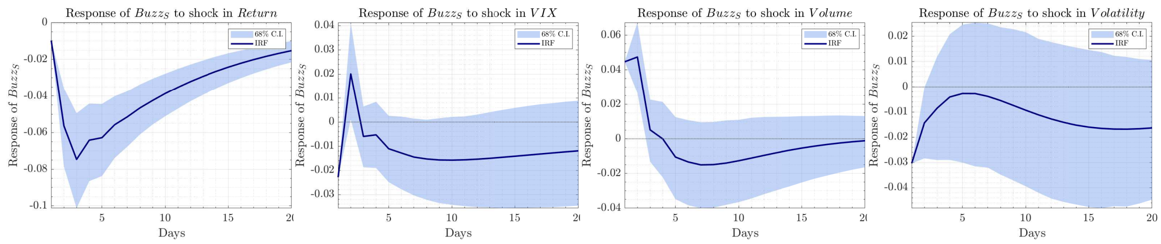


(a) r_t shock Sub 1 (b) V_t shock Sub 1 (c) vo_t shock Sub 1 (d) σ_t shock Sub 1

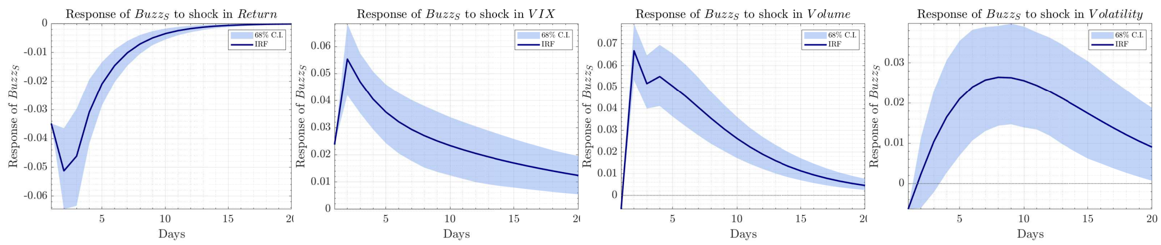


(e) r_t shock Sub 2 (f) V_t shock Sub 2 (g) vo_t shock Sub 2 (h) σ_t shock Sub 2

Figure A.5: $Buzz_N$ Responsce to Market Variable Shocks - Subsample Comparison. Panel(a) to (d) are IRFs for $Buzz_N$ responses from one standard deviation market variable shock in subsample 1, panel(e) to (h) are constructed in the same way but for subsample 2. All time-series are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.

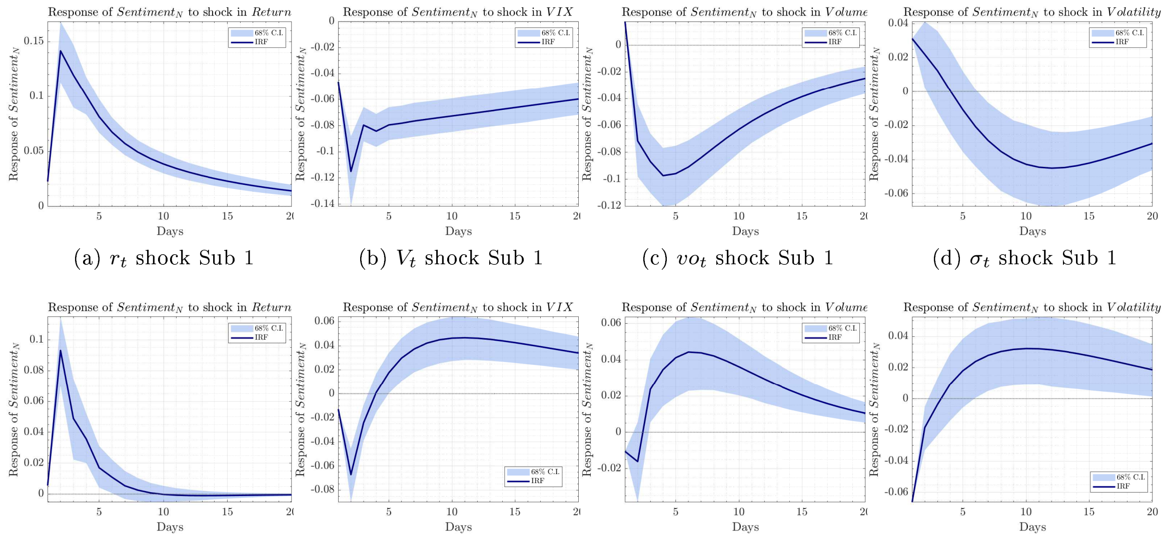


(a) r_t shock Sub 1 (b) V_t shock Sub 1 (c) vo_t shock Sub 1 (d) σ_t shock Sub 1



(e) r_t shock Sub 2 (f) V_t shock Sub 2 (g) vo_t shock Sub 2 (h) σ_t shock Sub 2

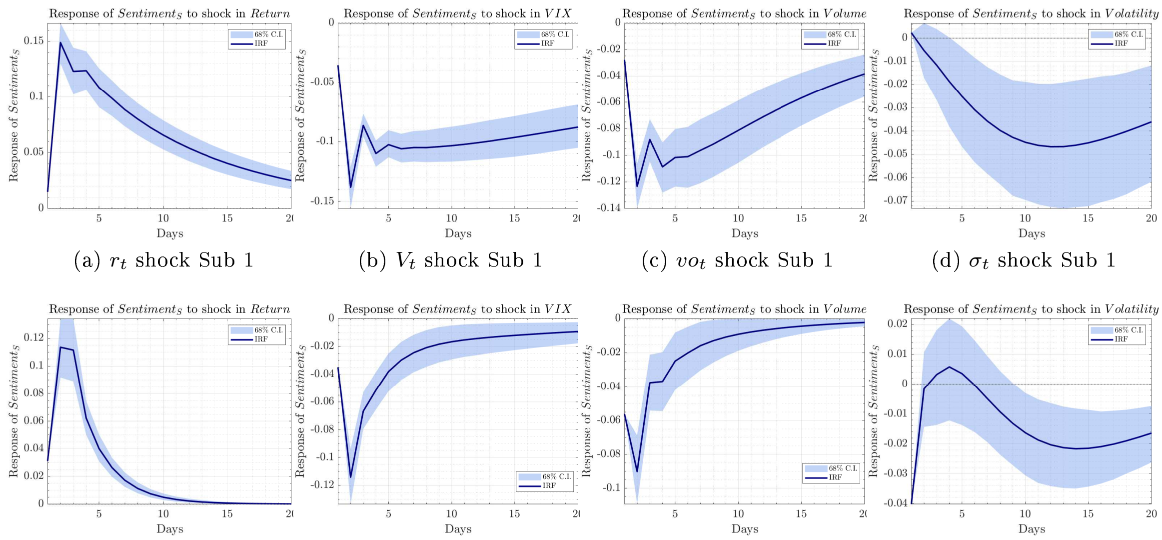
Figure A.6: $Buzz_S$ Responsce to Market Variable Shocks - Subsample Comparison. Panel(a) to (d) are IRFs for $Buzz_S$ responses from one standard deviation market variable shock in subsample 1, panel(e) to (h) are constructed in the same way but for subsample 2. All time-series are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.



(a) r_t shock Sub 1 (b) V_t shock Sub 1 (c) vo_t shock Sub 1 (d) σ_t shock Sub 1

(e) r_t shock Sub 2 (f) V_t shock Sub 2 (g) vo_t shock Sub 2 (h) σ_t shock Sub 2

Figure A.7: $Sent_N$ Responsce to Market Variable Shocks - Subsample Comparison. Panel(a) to (d) are IRFs for $Sent_N$ responses from one standard deviation market variable shock in subsample 1, panel(e) to (h) are constructed in the same way but for subsample 2. All time-series are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.



(a) r_t shock Sub 1 (b) V_t shock Sub 1 (c) vo_t shock Sub 1 (d) σ_t shock Sub 1

(e) r_t shock Sub 2 (f) V_t shock Sub 2 (g) vo_t shock Sub 2 (h) σ_t shock Sub 2

Figure A.8: $Sent_S$ Responsce to Market Variable Shocks - Subsample Comparison. Panel(a) to (d) are IRFs for $Sent_S$ responses from one standard deviation market variable shock in subsample 1, panel(e) to (h) are constructed in the same way but for subsample 2. All time-series are standardized to have 0 mean and variance equal to 1. Error bands are constructed at the 68% interval following Sims and Zha (1999). Subsample 1: 01/Jan2011-28/Feb/2014, subsample 2: 01/Mar/2014-30/Nov/2017.