Investor Sentiment Under Microscope*

Baoqing Gan†, Vitali Alexeev, and Danny Yeung

Finance Department, UTS Business School, Sydney, Australia

Preliminary Version: July 28, 2019
(please do not circulate)

Abstract

Using high-frequency investor sentiment metrics based on multiple news and social media outlets, this study investigates the intraday dynamics between sentiment and stock returns. Individual stocks and sentiment data for all Dow Jones Industry Average (DJIA) constituents are considered for the period from 2011 to 2017. We find that large sentiment tensions accumulated overnight have strong predictability on the next day opening returns, but it is asymmetric under negative and positive sentiment. Robustness checks for the first minute of trading, return autocorrelation, and returns in excess of those on the broad market index are conducted. We find that social media signals generate daily excess returns of 15 basis points, which is much higher than that based on news media singals. Overall, this paper contributes to the literature on overnight sentiment and intraday return patterns, which brings about new insights to the day and night return puzzle.

Keywords: overnight sentiment; opening prices; textual analysis; high-frequency; TRMI

JEL: G14, G40, G41

---

*We thank Thomson Reuters Financial and Risk for offering MarketPsych Indices (**TRMI**) as part of our research data; we are grateful to Ronald Bird in assisting with TRMI dataset application. This research is supported by an Australian Government Research Training Program Scholarship. We thank helpful comments from faculties attending Autumn 2019 UTS Finance Stage Assessment Presentations.

†Corresponding author: PhD student and casual academic at UTS Business School, Email: *baoqing.gan@uts.edu.au*; Phone: +61 2 9514 7787. This research constitutes a part of the corresponding author's PhD thesis. Baoqing Gan is grateful to the continuing support, guidance and encouragement from the supervisory panel: Vitali Alexeev *vitali.alexeev@uts.edu.au*, Danny Yeung *danny.yeung@uts.edu.au* and Ronald Bird *rgbird73@gmail.com*.

1 Introduction

The rise of mobile devices and social media is changing our everyday life. According to the Global Digital 2019 Reports, 3.26 billion people are using social media on mobile devices in January 2019, with a growth of 297 million new users - a year-on-year increase of more than 10 percent. Nowadays, the majority of firm-specific announcements are scheduled outside of trading hours (Birru, 2018),[1] with news articles and social media feeds arriving continuously, creating round-the-clock information flows (Kelley and Tetlock, 2013) to two types of investors: retail investors and professional investors. The former is uninformed and prefers to trade at or near the market open, while the latter is informed and prefers to trade during the rest of the day or near the close (Lou et al., 2019). A natural question that arises to this background is how and to what extent does the overnight investor sentiment in media impact on the next day returns?

Using textual analysis investor sentiment metrics of individual stocks that capture social media and news media feeds, we investigate the build-up of positive and negative overnight sentiment and their dynamics with the next day opening returns. We concentrate on contrasting the effects of social media sentiment and that of news media. Specifically, we aim to answer three research questions in this paper: (1) Does overnight investor sentiment predict opening stock returns the next trading day? (2) How do the ups and downs of daily stock returns affect the after-hour investor sentiment in media? (3) How do these patterns change when sentiment is based on social media rather than news media? The first question addresses the predictive property of sentiment, while the second question investigates the descriptive property of sentiment. The third question is the main perspective that leads our methodology and will be built in empirical tests that study the previous two questions. We consider all individual stock sentiment and prices of all Dow Jones Industry Average (DJIA)[2] constituents from 2011 to 2017. We choose DJIA stocks to mitigate the problem of unavailable observations in high frequency analysis, which is in line with the idea of "salient" stocks of Akhtar et al. (2012),

We conduct two sets of tests to study these questions, and proceed by several robustness checks. In the first group of analysis, we document the opening return patterns. We sort the build-up of overnight investor sentiment in social and news media respectively, and concentrate on the top and bottom decile of overnight sentiments. We observe their associations with next day returns by varying the corresponding return series windows, i.e. in the first half hour, first hour, and morning trading sessions. To control for possible market-wide announcement and events effects, we take difference the DJIA index return from individual stock returns. In order to avoid the influence from overnight returns (close-to-open), which is usually captured in the first minute of trading, we also conduct comparative analysis by including and excluding the first minute of trading. We reveal the following facts. First, influence on the opening return is different when overnight sentiment is based on social media instead of news. Second, overnight sentiment and next day opening return are strongly positively correlated. Aggregating from 9:30am, the correlation between strong positive and negative social media sentiment

[1] Some studies have corroborated this result. In the Jiang et al. (2012) research sample, over 95% announcements are not in the regular trading periods. Similarly, Bagnoli et al. (2005) reports that the ratio of earnings announcements happened in trading hours have slid down from 67% in the 1990s to only 27% for the 2000 to 2005 periods. Michaely et al. (2013) presents that only 5% corporate earnings announcements occurred during trading hours in the period from 2006 to 2009. Similarly, Bradley et al. (2014) also documents that most of earnings announcements and analyst upgrades take place outside of trading hours.

[2] A list of acronyms is provided in Table A.1 in page 23 of the appendix.

and the first half hour cumulative excess return (CAR) is 0.3926 (first hour 0.3759, morning 0.3997), while such correlation between news media sentiment and first half hour CAR is 0.5684 (first hour 0.5147, morning 0.5464). However, when we exclude the first minute and aggregate return from 9:31am, the correlation between social media sentiment and CAR drops substantially, with 0.1268 for the first half hour, 0.0758 for the first hour, and 0.0969 for the morning session. Similar plunges are also found in the news media sentiment effect: correlation coefficients between first half hour, first hour, and morning CAR with top/bottom cumulative sentiment equals to 0.1368, 0.0304, and 0.1191 respectively. Lastly, the top and bottom decile overnight sentiment signals potential outperforming strategies. The cross-sectional average profit when longing high and shorting low social media sentiment ranges from 19.25 to 21.26 basis points (b.p.), and such profit when taking news media sentiment as signal ranges from 14.20 to 14.51 b.p.. Yet, these mispricing opportunities also diminish quickly when returns are aggregated from 9:31am rather than 9:30am. In this way, we also provide evidence that overnight return is a suitable proxy of individual company investor sentiment, which is consistent with Aboody et al. (2018).

In the second group of analysis, we document the after-hour investor sentiment patterns. Sorting the daytime cumulative excess returns for each sampling stock and focusing on the top/bottom deciles, we display the associated after-hour investor sentiment patterns in social media and news respectively. We find that after-hour investor sentiment are significantly positively related to the daytime stock performance. Correlation between after-hour social media sentiment and the top/bottom decile performance days is 0.4949, and that of news media sentiment is 0.7086. According to bootstrap simulations, these results are statistically more significant than the impact from overnight sentiment on opening returns.

Based on these findings, we conduct robustness checks by performing double-sorting procedures to control for previous day return effect (daily return autocorrelation) and examine the conditional opening return patterns. We find consistent results that the next day cumulative excess returns (from 9:31am) are positively linked to the overnight sentiment in social (correlation of 0.2844) and news media (correlation of 0.0905), which greatly resolves the endogneity problem in the previous two groups of tests. Controlling for the previous day performance, a strategy longing high social media sentiment and shorting low social media sentiment days generates an average daily cumulative excess returns 15.06 b.p. We also observe that the short-leg (negative sentiment side) on average produces higher profits than the long-leg (positive sentiment side) does.

We contribute to the literature in three ways. Firstly, we apply granular individual stock specific sentiment measures instead of general market based sentiment metrics. Since Thaler (1987) promoted investor sentiment as an explain to the day and night return puzzle (Cooper et al. (2008) and Branch and Ma (2012)), prior studies (e.g. Berkman et al. (2012) and Aboody et al. (2018) among others) investigate the sentiment effect using general market sentiment measures such as Baker & Wurgler's investor sentiment index (BW) or the the Index of Consumer Sentiment (ICS) provided by the University of Michigan's regular survey. Research that applies high-frequency data and individual stock specific investor sentiment measures, however, is relatively rare. Therefore, we contribute to the literature such as Sun et al. (2016), Renault (2017) and Behrendt and Schmidt (2018), which concentrates on novel high-frequency investor sentiment proxies. Secondly, we provide new insights to the asymmetric influence from investor sentiment on the opening returns. Recent empirical results relating to this asymmetry are somewhat mixed. For instance, Barber et al. (2008) argues that there are higher uninformed buying pressures than selling at the open, due to the limited attention and the short-selling constraints on retail investors. On the other hand, Stambaugh et al. (2014) suggests that such effect should be negative because, under high investor sentiment environment, the opening hour short-leg portfolio profits are higher than the long-leg. We add value to this line of research by providing detailed comparison between the positive and negative sentiment induced excess returns at the open. Last but not least, though literature on textual analysis investor sentiment is growing rapidly in recent years, few studies bifurcates the social media effect from news media. To the best of our knowledge, Jiao et al. (2016), among others, is the only study that emphases such distinction. We differ from Jiao et al. (2016) in two ways, however. First, our study uses more granular data at 1-minute frequency while Jiao et al. (2016) uses sentiment data at a lower frequency (monthly). Second, Jiao et al. (2016) is focusing on using the volume of media activities (*Buzz*) to explain the volatility and volume of stock market, we focus on the first moment of stock prices: return, which is less commonly investigated.

This paper proceeds as follow: Section 2 reviews literature and formally propose the hypotheses. Section 3 describes in detail about sample data and methodology. Main results are provided in Section 4. In Section 5, we proceed with robustness tests and discussions about what drives our main findings. Section 6 concludes the paper.

2 Literature and Hypothesis

2.1 Day and night return puzzles

It is documented extensively in empirical finance literature that the intraday mean stock return and volatility are U-shaped.[3] Using time-sequencing stock transaction tape data between December 1981 and January 1983 from *Francis Emory Fitch, Inc.* , Harris (1986) decomposes the close-to-close daily returns of the New York Stock Exchange (NYSE) into trading- and non-trading components to investigate the day-of-the-week effect. Interestingly, after further breaking down the day (open-to-close) returns into a series of 15-minute intraday returns, Harris (1986) identifies that the mean intraday returns at the beginning and end of trading hours are five- to ten-folds higher in absolute value than cumulative returns in the middle of the day. As a pioneering work, this study shows that the significant intraday return differences accrue during the first 45 minutes after the market opens. Similar patterns are also observed in Jain and Joh (1988), which uses hourly intraday data on NYSE stocks and Standard and Poor's 500 index (S&P 500) to examine the inter-dependence between volume and return within trading hours, with a longer period (5 years) than Harris (1986) does.

This phenomenon has triggered researches' interest in testing whether the U-shaped return pattern and the day-vs-night return differences occur in other financial markets, and what factors lead to this pattern. Cooper et al. (2008) provides evidence that equity returns during daytime trading periods (open-to-close) are smaller than overnight returns (close-to-open) and claims that this finding is consistent across multiple US equity markets. Their exploration shows that factors like risk, the schedule and magnitudes of earnings surprises, the presence of ECNs and decimalisation, return autocorrela-

[3] Hong and Wang (2000) summarises 5 empirical patterns on the intraday return, volatility, and trading volume, and reviews their relevant literature in detail.

tions, and liquidity, only provide partial explanation to the overnight and daytime return differences. Kelly and Clark (2011) compares the day and night returns for a group of exchange-traded funds (ETFs). Using risk-adjusted excessive returns and the value-weighted average prices (VWAPs) in the first and last 5 trading minutes, Kelly and Clark (2011) finds similar results as documented in Cooper et al. (2008) that the overnight return is significantly higher than the daytime return. Branch and Ma (2012) also finds negative correlation between overnight return and the subsequent daytime return, and discusses the possible causes to this intraday return anomaly. They focus on market microstructure factors, which include the trading behaviours of market makers and sophisticated investors to avoid being adversely selected, the bid-ask bounce effects between the prior day closing and the new opening prices, as well as the collective emotions of the unsophisticated investors during the non-trading periods.

Thaler (1987) summarises empirical evidence as mentioned above, and points out three explanatory hypothesis in terms of institutional investors: (1) The inflow and outflow of funds to the market; (2) The "window-dressing" behaviour of portfolio managers; and (3) The timing of the arrival of good and bad news. These three hypotheses, however, do not account for any effects from individual investors, which comprises a larger fraction of market participants. Accordingly, Thaler (1987) further argues that variation in the mood of market participants should also be considered. A new strand of literature, hence, explains this day and night return puzzle from the behavioural finance perspective. In particular, this line of research concentrates on analysing effects from retail investors' sentiment formed by receiving attention-grabbing news. For example, using squared return and net buying volume at the opening on the prior trading day as two proxies of retail investors' attention, Berkman et al. (2012) proposes a mechanism that explains the opening price formation process and disentangle the intraday trading hour return reversal pattern. They extend hypothesis of Barber et al. (2008) and show that high attention days are followed by retail investors' high net buying behaviour at the commencement of the next day. This net retail buying pressure pushes the opening hour price to deviate from the rest of the trading hours of the day. They also find that stocks that are difficult-to-value and hard-to-arbitrage are most significantly affected by this intraday price pattern. Berkman et al. (2012) further conducted investigations on whether the magnitudes of this day-vs-night return differences is exacerbated during high general investor sentiment environment. They find that under high market sentiment condition, such mean trading day reversal is more than twice the size of the effective half spread. Based on the assumption of Berkman et al. (2012) that retail investors are the mostly affected market participants by the attention-driven news during non-trading hours, Aboody et al. (2018) argues that overnight return could be a suitable proxy of firm-specific investor sentiment, and shows that stocks with high (low) overnight returns will underperform (overperform) over the long run, which is in line with the hypothesis of temporary sentiment-induced mispricing. They also provide evidence that, the short-term persistence of overnight return is in line with the increased demand on stocks by sentiment-driven investors. And they also corroborate that this phenomenon is more prominent for the difficult-to-arbitrage and hard-to-value firms.

2.2 Asymmetric sentiment effect on return

Early studies on investor sentiment and return predictability show that there is an asymmetric effect between positive and negative sentiment via media coverage that catches investors' attention. Some document a “positive effect”. For example, Chen et al. (2004) reveals results that the additions to and deletions from the S&P 500 stock index have asymmetric price response effects: the negative effect of deletions is smaller than positive effect of additions, or even close to non-exist. They argue that such asymmetric response is better explained by the extent of investor awareness after the addition and deletion events. Barber and Odean (2007) points out that attention is a scarce resource. They propose and test the hypothesis that individual investors incline to trade on attention-grabbing stocks, and there is more buying than selling for such stocks. Contrary to theoretical models such as Grossman and Stiglitz (1980) and Kyle (1985), which assume that retail investors take buying and selling activities symmetrically, Barber and Odean (2007) proves that retail investors will allocate more resources in buying attention-grabbing stocks than selling them, due to the pre-requisites of selling: one needs to hold the asset before selling it, or, at least, short-selling should be achievable, which may not always be viable to retail investors. Palomino et al. (2009) argues that due to limited information processing abilities of individuals, the way information is processed may depend on its relative salience (media coverage). They find evidence that investor sentiment causes an asymmetric share price reactions toward wins and loses for London Stock Exchange (LSE) listed soccer club companies: the abnormal returns induced by positive (winning) sentiment is higher than the market reaction to a loss induced by negative emotions.

Others provide evidence of “negative effect” at both market level and cross-sectional stocks level. For instance, using the Index of Consumer Sentiment (ICS) from Thomson Reuters/University of Michigan from January 1991 to August 2010, Akhtar et al. (2012) finds that both the stock index and futures returns have stronger negative reactions to negative sentiment “surprise” than their positive counterparts. They attribute such negative effect to the “salience” of stocks - stocks that are relatively large, frequently covered in media, and followed by analysts, resulting from the availability heuristic of Tversky and Kahneman (1973). Specifically, they show that the returns of Dow Jones (DJ) index and its futures exhibit a significant negative announcement day effect when the ICS announcement is lower than the previous month. And the magnitudes of DJ’s effect are larger than that of S&P index and its corresponding futures. Stambaugh et al. (2012) investigates the relationship between investor sentiment and the cross-sectional stock returns, and finds consistent predictability of investor sentiment across 11 market anomalies. Specifically, they reveal evidence that supports three hypothesis: first, there is a positive long-short profit following higher sentiment; second, increases in sentiment lead to higher short-leg profit; and third, current sentiment exhibits no relations to future long-leg returns. To relieve the concern that investor sentiment might be a spurious regressor in the Stambaugh et al. (2012) study, Stambaugh et al. (2014) strengthens the previous research results by replacing the main variable into more than 200 million simulated regressors. Since none of the simulated regressors performs as strongly as investor sentiment does in terms of the predictability consistency across 11 anomalies, they claim that investor sentiment is, indeed, an effective variable to predict the cross-sectional stock returns.

2.3 Sentiment from novel high-frequency data sources

Based on the aforementioned attention-driven sentiment hypothesis, extensive of studies have been carried out to examine the effects from news-induced sentiment on the intraday stock return patterns. Facilitated by the availability of novel data source with granular frequency, as well as the improvement in machine learning and textual analysis, current researches are able to investigate sentiment from different kinds of sources and try to explain their effects on stock returns at intraday (high-frequency) level.

Using viewship data from *Nielson Media Research* to quantify investor attention that is caught by a stock recommendation TV program named *Mad Money*, Engelberg et al. (2012) finds result that higher attention to the show leads to higher overnight returns. Moreover, assisted by short-sale lending data, they directly test the proposition of Barber and Odean (2007) that the attention-driven sentiment from retail investors is asymmetric toward buying than selling activities. Controlling for impacts from other news announcement, they prove that the the buying recommendation from *Mad Money* has a larger effects on the overnight returns than the selling recommendation does.

Applying high-frequency textual analysis sentiment data that captures news and social media content from *Thomson Reuters MarketPsych Indices* (TRMI), Sun et al. (2016) tests the within-day half-hour return predictability from media sentiment. They find that the intraday S&P 500 index returns are predictable by the lagged changes of half-hour investor sentiment, controlling for day-of-the-week effect. They also document such predictability in the stock or bond index ETF markets. Sun et al. (2016) claims that this within-day sentiment effect is different from the intraday momentum effect (Gao et al., 2018), as the former usually persists for at least two hours while the latter only presents for the last half hour.

Although various empirical results as mentioned above have proved the return predictability of investor sentiment using daily data or at a lower-frequency frequency, whether sentiment from social media platforms, such as Twitter and StockTwits, at intraday level remains to be inconclusive. For example, Renault (2017) develops a novel approach of lexicon of words for messages on the microblogging platform StockTwits, and finds that investor sentiment changes in the first-half trading hour help predict the last-half hour return of the S&P 500 index ETF. They also documents that such short-term price deviations induced by investor sentiment in StockTwits is followed by a price reversal on the next trading day. On the contrary, Behrendt and Schmidt (2018) examines the relationship between Twitter sentiment, as well as quantity of Tweets, and absolute 5-minute returns of the constituents of the DJIA. They argue that the economic effect of high-frequency Twitter sentiment and activity is at a negligible magnitude, though there are some statistically significant co-movements between the Tweets information and the intraday return volatility.

2.4 Empirical Hypotheses

To help narrow the literature gap with respect to the inconclusive and contradictory empirical evidence as mentioned above, we propose to test the following hypotheses in this study:

Hypothesis 1: *Overnight sentiment from social media and news media are positively associated with the next day opening returns.*

Hypothesis 2: *The correlation between social media sentiment and next day opening returns is different from the correlation between news media sentiment and next day opening returns.*

Hypothesis 3: *Negative overnight investor sentiment in media has stronger effects on the next day returns than positive overnight sentiment does.*

3 Data and Methodology

3.1 Overnight Sentiment Data

To mitigate possible sampling bias from missing observations in high-frequency analysis, we choose the Dow Jones Industry Average (DJIA) constituents to conduct this study, which is consistent with the definition of "sailent" stocks of Akhtar et al. (2012). Akhtar et al. (2012) argues that stocks that are more "salient" to investors are more sensitive to sentiment, which are not necessarily the sentiment-prone stocks. Sentiment-prone stocks are small, young, unprofitable with high growth, highly volatile, and non-dividend paying, as characterised in Barber and Odean (2007). Salient stocks, however, are securities that are more prominent, or "iconic" in the market. Good candidates for salient stocks are large stocks with more discussions in the press and followed by more analysts (Akhtar et al., 2012).

Our company specific investor sentiment data comes from Thomson Reuters MarketPsych Indices (TRMI), a proprietary dataset that scrapes and scores texts from various news press and social media via textual analysis algorithms, and generates both quantities and emotion scores.[4] To suit the purpose of contrasting impacts from different media outlets, we use sentiment scores based on **social** media and **news** media respectively at 1 minute frequency - the most granular data available by TRMI. Our sample period is from 1 January 2011 to 30 November 2017, which avoids possible confounding effects from the Global Financial Crisis (GFC) from 2008 to 2010 but covers a period when social media is developing swiftly.

Table A.3 in the appendix provides number of observations for the 35 DJIA sampling stocks.[5] $Buzz_S$ ($Buzz_N$) and $Sent_S$ ($Sent_N$) represent the volumes of postings in social (news) media and the net emotional scores in social (news) media respectively.[6] To better reflect how "salient" each stock is, we convert the TRMI observations of each stock into average daily number of observations, and plot them in Figure A.1 and Figure A.2 in the appendix. In Figure A.1, blue bars indicate the average daily volumes of postings in social media, and orange bars represent average daily number of reports in news wires. Similarly, in Figure A.2, blue bars are the average daily number of postings in social media that conveys positive and negative emotions, while orange bars are those in the news media. We observe that our sample stocks display a remarkably different patterns in terms of "salience": some are more salient in social media (higher blue bars) whereas others are covered more in news reports (higher orange bars). Moreover, technology stocks generally have more coverage in both social and news media than stocks from other sectors. For instance, stocks like Apple, Microsoft, Cisco, and Intel that rank in the left of Figure, contain media coverage that are at least 40 times of the least covered stock: e.g. Travelers.[7] A comparison between Table A.1 and Table A.2 also reveals that, the "salience" in sentiment series ($Sent$) are generally consistent with buzz series ($Buzz$). In this research,

[4]This data is provided by Thomson Reuters Financial and Risk Team as part of TRMI product. Markets and security coverage of TRMI include: over 12,000 companies, 36 commodities and energy subjects, 187 countries, 62 sovereign markets and 45 currencies since 1998, and more than 150 cryptocurrencies since 2009. A detailed summary of this dataset and description is provided in *Thomson Reuters MarketPsych Indices 2.2 User Guide*, 23 March 2016, Document Version 1.0.

[5]Delisted DJIA stocks are also incorporated for the initial consideration because they fit the definition of salient stocks.

[6]A list of variable notations and their corresponding definition is provided in Table A.2 in page 24 of the appendix.

[7]Daily $Buzz_S$ of Intel equals 104.4, daily $Buzz_S$ of Travelers is only 2.5; daily $Buzz_N$ of Intel equals 92, and daily $Buzz_N$ of Travelers is only 2.3. Daily $Sent_S$ of Intel is 88.8, daily $Sent_S$ of Travelers is 1.8; daily $Sent_N$ of Intel is 81, and daily $Sent_N$ of Travelers is only 2.

we concentrate on the *Sent* series because they are volume-weighted net scores of positive and negative emotions in news and social media.[8]

3.2 Stock Price Data

The Dow Jones Industry Average (DJIA) index and constituent stocks data are obtained from Thomson Reuters DataScope. We abstract the 1-minute closing, ask and bid prices from 2011 to 2017. Each missing observation is replaced with its previous value, and if the missing value appears at the beginning of total sampling period, we replace it with the next available observation. For any individual stock i in DJIA, the return of day t at time j ($R_{i,t,j}$) is calculated as logarithm return using the mid-price ($P^m_{i,t}$) in order to avoid possible market microstructure confounding impacts: $R_{i,t,j} = ln(\frac{P^m_{i,t,j}}{P^m_{i,t,j-1}})$, where $P^m_{i,t,j} = \frac{1}{2}(A_{i,t,j} + B_{i,t,j})$ is the average of the Bid ($B_{i,t,j}$) and Ask ($A_{i,t,j}$) prices.[9] Mid-price return series for the broad market index (RIC: .DJI) $R_{m,t,j}$ are computed in similar manner.

3.3 Data Aggregating Method

To deal with the irregularity and non-synchronisation problem, we first fill all missing observations along the 24 hour window over our whole sampling period.[10] For sentiment series, which range between -1 and 1 and center at 0, we replace the missing observations with 0. For the return series, we replace the missing value using the next available observation.

We regard each day market open as similar to an "event", and use the MacKinlay (1997) event study methodology in an intraday context.[11] Our event window is constructed as a $(T_1 + 1 + T_2)$ minutes interval: the T_1 minutes indicate overnight period $[I_{(-T_1)}, I]$, for example [16:00pm day (t-1), 9:29am day (t)]. An one minute at 9:30am is considered analogue to the announcement time. The interval $[I, I_{T_2}]$ represent "opening hour" period, for example [9:31am, 10:30am] in day (t). Let $\hat{\mathbf{AR}}_i$ denote a $((T_1 + 1 + T_2) \times i)$ vector of abnormal returns for each stock i of the DJIA between time $I_{(-T_1)}$ and I_{T_2}.

Let $p = (T_1, (T_1 - 1), \ldots, 2, 1)$ denotes the overnight period, $Sent_S$ denotes social media sentiment score at any point in time, $Sent_N$ represent news media sentiment score at any point in time. Because we are interested in the build-up of sentiment over a specific time period, the pre-filling process does not compromise accuracy of the sentiment series. We therefore, define and denote the overnight cumulative social media sentiment as $CSent_S = \sum_{p=T_1}^{1} Sent_S$, and the overnight cumulative social

[8]See Thomson Reuters MarketPsych Indices 2.2 User Guide, 23 March 2016, Document Version 1.0, Chapter 13, page 32: *'...all emotional measures are "buzz-weighted" indices'.*

[9]In an unreported test, we also computed return without controlling for the market microstructure effect using "last" prices from DataScope: $R^l_{i,t,j} = ln(\frac{P^l_{i,t,j}}{P^l_{i,t,j-1}})$. The results are similar and upon request. A full list of variable names and their definitions can be found in Table A.2 of the appendix in page 24.

[10]The mid-price return series and the TRMI sentiment series are irregular at 1 minute frequency. Whereas returns are generally continuous over trading hours, TRMI sentiment observations only pop up when there is someone posting in the news wires or social media. Returns are concentrated during trading hours from 9:30 to 16:00, while TRMI series show up irregularly round-the-clock.

[11]The classic event study methodology is used widely in measuring market react to certain type of corporate events, such as: the earnings announcement, merger and acquisition, stock split for individual stocks, or macroeconomic announcement events, for example, sovereign debt rating downgrade, the federal fund rates change, etc. We do not actually research on any specific announcements in this paper, but merely using this approach to run a quasi-natural experiment cross-sectionally.

media sentiment:$CSent_N = \sum_{p=T_1}^{1} Sent_N$.

Let $t = (1, 2, \ldots, T_2)$ denotes the opening hour period, we aggregate the excess returns during the opening hour interval $[I, T_{T_2}]$ for each stock i and arrive at the cumulative excess returns: $\hat{CAR} = \sum_{t=1}^{T_2}(R_{i,t} - R_{m,t})$ In this way, we have three 1-minute continuous series: return $(CAR_{i,t,j})$, social media sentiment $(CSent_{S,t,j})$ and news media sentiment $(CSent_{N,s,j})$, with $j = 3,637,440$.[12] Next, we align sentiment series with return, by reshaping the three 3,637,400 1-minute series into 2,526 days, with non-missing observations. Lastly, we remove days that contains more than 95% of zero-returns. We prevent from conducting pooled average analysis across stocks because of the computational limit from this large number data processing. Instead, we report the individual stock specific results of sorting $CSent_S$ ($CSent_N$) and the conditional CAR in Section 4.

Table A.4 in the appendix summarises total "events" number, their distribution, and the "events" that have been removed because of too many zero-returns for our sample stocks, as well as these of the DJIA index. Panel A describes social media sentiment and Panel B describes news media sentiment. As shown in Table A.4, Kraft (RIC: KFT.OQ) has too sparse events left for both social media and news-based sentiment after our reshaping procedure, perhaps subject to the latest Merger and Acquisitions and the updates of the Ticker. We therefore remove Kraft from our sample for further studies. The normal sampling stocks contain 1,741 event days. Whereas most stocks have an evenly distributed cumulative sentiment across the 1,741 days (e.g. AAPL.OQ, IBM.N and JPM.N), there are other stocks whose cumulative sentiment are mainly concentrated in the highest and lowest percentiles (e.g. TRV.N and UNH.N). The differentiation between social media sentiment and news media sentiment facilitate to study our second research question: what is the difference between social media and news media effects? To investigate the third research question: what effect does the daytime return have on the after hour media, we follow the above "quasi-event" analogy by replacing our "event" into the market closure instead of market open, and observe the changes in cumulative sentiment.

4 Findings

4.1 Opening Return Patterns

Using IBM as an example, we demonstrate how we conduct the analysis in Figure 1. The top figure indicates decile sorted cumulative **social** media sentiment from the closing of previous day to the opening of next day ($CSent_S = \sum_{t_0=16:01,(t-1)}^{t_1=9:29,t}[Sent_S]$), as well as the corresponding next day cumulative excess returns ($CAR = \sum_{t_2=9:30}^{t_3=16:00}[R_{IBM,t} - R_{m,t}]$). Left-hand side of the figure shows 10 equal-sized bin of $CSent_S$ series, with the top thick blue curve indicating the highest 10% $CSent_S$ and the bottom thick red curve representing the lowest 10% $CSent_S$. The dashed black curve shows the average $CSent_S$ of the 10 equal-sized bins. The right-hand side of top panel in Figure 1 shows the CARs conditional on the overnight cumulative sentiment decile, as mapped through the same colour of lines. The dashed black curve is the average cumulative excess return ($ACAR$). The three layers of grey shaded area are the 90%, 95% and 99% confidence bands constructed by performing bootstrap re-sampling procedure with replacement and repeating 2,000 times, using sample size of 174

[12] Our sample is of 2,526 days from 1 January 2011 to 30 November 2017, which corresponds to $2,526 \times 24 \times 60 = 3,637,440$ minutes.

- the sample size of each equal-sized decile events for IBM. The bootstrap randomisation generates the unconditional CAR. The bottom panel of Figure 1 represents those of the **news** media sentiment.

To relieve the concern that the first minute of trading is equal to overnight return (close-to-open), which may interfere with the return at opening, we conduct the same analysis by excluding the first minute return (accumulating excess returns from 9:31am to 16:00pm). The results for IBM is provided in Figure A.3 of the appendix (page 29).

Both social media and news media panels in Figure 1 show that the conditional cumulative excess returns induced by the top and bottom deciles of overnight cumulative sentiment are positively associated with sentiment, and they are statistically significant at 99% confidence level, varying outside of the bootstrap simulated significance band. The cumulative excess returns corresponding to other decile bins are, however, not statistically significant as shown deviating in the greyed area. Eliminating the overnight return, in Figure A.3, we observe that CAR corresponding to the top decile $CSent_S$ (thick blue lines in the right-hand side, top panel) continues trending up in the morning, and fluctuates to the end of trading time, statistically significant at 90% level, while the bottom decile $CSent_S$ related CAR is not significant. The bottom panel of Figure A.3 shows that the most negative news media emotion associated returns (red thick lines) becomes insignificant, too, after eliminating the first minute return. The most positive emotions in news media, however, is related to CAR trending upward in the morning and flattens out the next day, at 99% significance level - more prominent than the social media sentiment does. We conduct this analysis throughout all sample stocks and summarise the results in Table 1. We also tested different windows: the first half hour, the first hour, and morning sessions of trading, the results are similar and do not affect our conclusion.[13]

Several interesting results are uncovered from Table 1. Firstly, the highest and lowest overnight sentiment is positively associated with opening cumulative excess return the next day. And the such impact from news media is more salient than from social media, as shown by higher correlation coefficient. Table 1 shows that the correlation between extreme $CSent_S$ (top and bottom decile overnight sentiment based on social media) and the corresponding $CAR0_S$ (cumulative excess return aggregated from 9:30am to 10:00am) is 0.3926, while the correlation between extreme $CSent_N$ (top and bottom decile overnight sentiment based on news media) and the corresponding first half hour $CAR0_N$ is 0.5684. Excluding the first minute, however, the correlation between extreme $CSent_S$ and the corresponding $CAR1$ (cumulative excess return aggregated from 9:31am to 10:00am) is only 0.1268, whereas the relationship between $CSent_N$ and the corresponding $CAR1$ is only 0.1386. The differences between positive sentiment and negative sentiment induced CARs, as shown in columns (3), (5), (8) and (10) in Table 1 create outperforming opportunities. For instance, the average profit across all sample stocks when one could long the H10% overnight sentiment and short the L10% overnight sentiment in social media, includes 9:30am, equals to 21.26 basis points (b.p.) in the first half hour, while such profit signaled from news media amounts to 14.41 b.p.. When CAR is considered even without the first minute, the profit is still available (3.41 b.p signaled by social media sentiment, and 1.54 b.p. signaled by news media sentiment).

Secondly, the relationship between overnight sentiment and opening return is most prominent in technology stocks and stocks with higher media coverage. For example, p-values of Apple, HP, and IBM

[13]Table A.5 in page 30 and Table A.6 in the appendix in page 32 provides similar results when we aggregate returns in the first hour and the morning session rather than the first half hour.

- stocks located in the left side of Figure A.2 with higher media coverage, are statistically significant at 95% level for both social media and news media overnight sentiment, while the p-values of stocks located in the right side of Figure A.2 are insignificant. This finding is consistent with Sul et al. (2016) that the emotional sentiment about a firm's stock with larger number of followers and spreading rapidly through social media is more likely to be incorporated quickly into stock prices.

Thirdly, overnight sentiment is quickly impounded into next day's opening prices. Focusing on the magnitudes of $CAR0$ and $CAR1$ in Table 1, and taking IBM as an example, the most negative 10% overnight sentiment from **social** media is linked with -44.82 basis point (b.p.) of cumulative excess return in the first half hour the next day, which is significant at 99% level. Eliminating the first minute, the cumulative excess return in the first half hour is only -0.65 b.p. and insignificant at 90% level. The difference between -44.82 b.p. and -0.65 b.p. is the overnight return. Moreover, the most positive 10% overnight sentiment form social media is associated with 18.59 b.p. cumulative excess return in the first half hour the next day, and is significant at 99% level, while excluding the first minute, such half-hour cumulative excess return equals 10.36 and is also significant at 99% level. Similarly, the first half hour CAR aggregated from 9:30am conditional on the top 10% negative overnight sentiment based on **news** media is -37.91 b.p. with 99% significance level, which revert to positive 0.21 b.p. if we aggregate the CAR from 9:31am. And the first half hour CAR accumulated from 9:30am conditional on the top 10% positive overnight sentiment based on news media equals to 22.00 b.p., which shrinks to 11.16 b.p.if we ignore the first minute. This result directly support Aboody et al. (2018) that the overnight return is an appropriate proxy for overnight sentiment of individual stocks.

Lastly, the overnight sentiment effect on opening return is slightly negatively skewed. In Column (2) of Table 1, there are 19 out of 34 sampling stocks that contain higher absolute value of $CAR0$ associated with the lowest 10% $CSent_S$ than the absolute value of $CAR0$ associated with the highest 10% $CSent_S$. Similarly, in Column (5) of Table 1, there are 20 out of 34 sampling stocks that contains higher absolute value of $CAR0$ corresponding to the lowest 10% $CSent_N$ than the absolute value of $CAR0$ based on the highest 10% $CSent_N$. Eliminating the first minute, there are 18 stocks (19 stocks) show higher "negative" side $CAR1$ than "positive" side $CAR1$ with respect to $CSent_S$ ($CSent_N$).

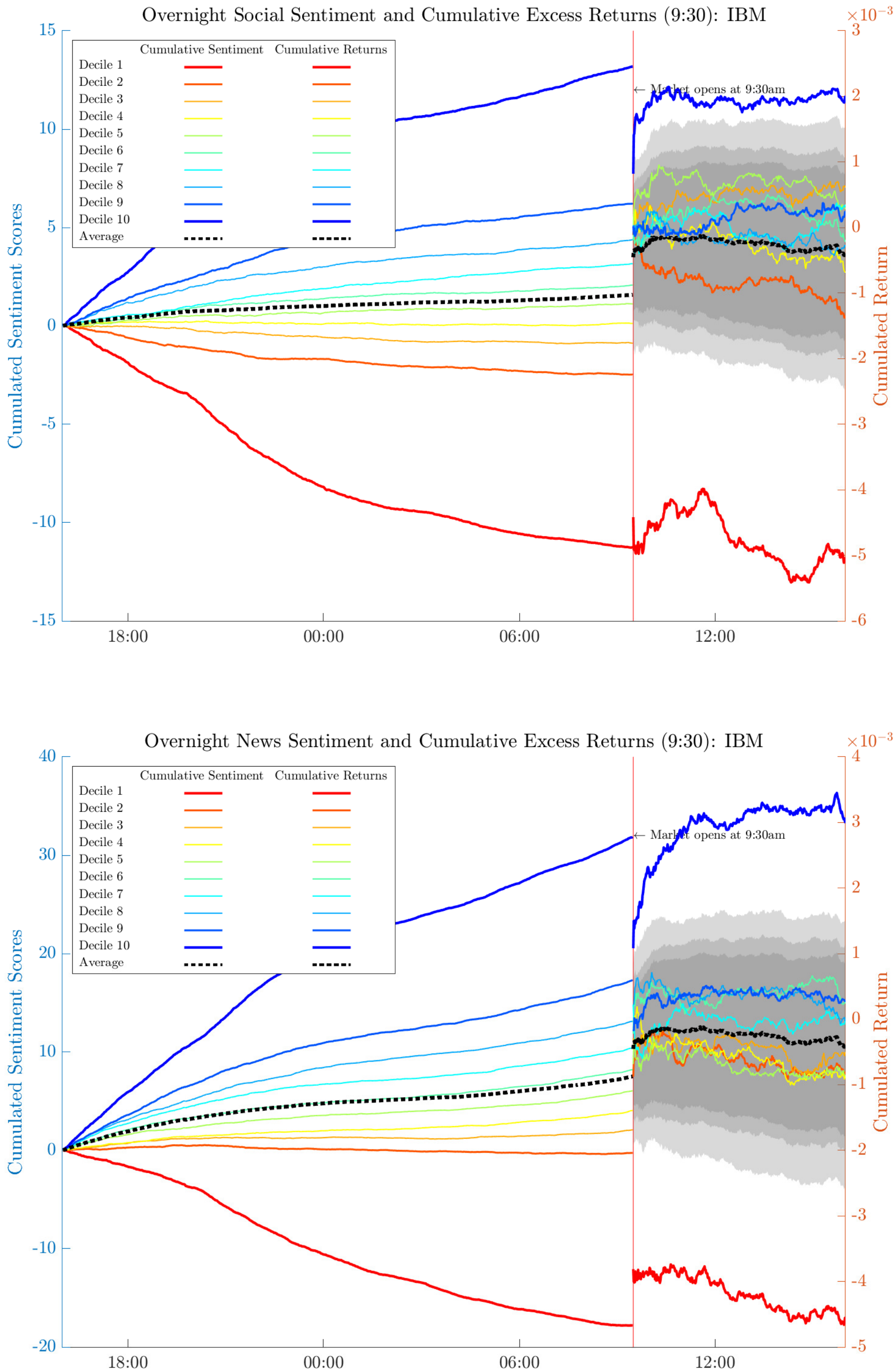


Figure 1: Example: IBM overnight media sentiment and next day cumulative returns (9:30am) This plots illustrate the equally sorted cumulative social media sentiment (top panel) and news sentiment(bottom panel) from previous day closing (4:00pm in day *t-1*) to market open (9:29am in day *t*), and the associated next trading day cumulative returns, accumulated from 9:30am to 4:00pm on day *t*. Sample period: 2011/01/01-2017/11/30, and the data is at 1-minute frequency. Left-hand side curves represent cumulative sentiment, and the corresponding colored curves at the right-hand side are the cumulative returns next day. The grey shaded confidence bands at 90%, 95% and 99% significance levels are constructed by performing bootstrap simulation using sample size of 174 (equal number of original sample size in each decile) with replacement and repeat 2,000 times.

Table 1: Extreme Overnight Sentiment and First Half Hour Cumulative Excess Return. This table reports the top and bottom decile cumulative sentiment of each sample stock from previous day market close to current day market open, and their corresponding cumulative excess returns during the first half hour of trading. Column (1) $\mathbf{CSent_S}$ and Column (6) $\mathbf{CSent_N}$ are cumulative sentiment of social and news media respectively, aggregating from 16:01pm to the 9:29am. Columns (2) $\mathbf{CAR0_S}$ and (7) $\mathbf{CAR0_N}$ are the cumulative excess returns aggregated from 9:30am to 10:00am as measured in basis points (b.p.), using the 1 minute mid-price log returns for each sample stock and subtracting mid-price log returns of the DJIA index. Columns (4) $\mathbf{CAR1_S}$ and (9) $\mathbf{CAR1_N}$ are the corresponding cumulative excess returns of each stock aggregated from 9:31am to 10:00am in basis point (b.p.). H10% and L10% represent the average value of the top and bottom decile cumulative sentiment. The correlation coefficient between strong social media sentiment and $CAR0_S$, including the first minute (the correlation between columns (1) and (2)), is 0.3926. The correlation coefficient between strong overnight news sentiment and $CAR0_S$, including the first minute (columns (6) and (7)), is 0.5684. Eliminating the first minute (9:30am), the correlation between strong social media sentiment and the first half hour CAR (column (1) and (4)) is only 0.1268, while the correlation between extreme news sentiment and the corresponding first half hour CAR (columns (6) and (9)) is 0.1386. Columns (3), (5), (8) and (10) "H-L" are the differences if we long highest sentiment CAR and short low sentiment CAR of each stock, expressed in columns (2), (4), (7), and (9). "Profit" indicates the average CAR differences across sample stocks when one longs the high sentiment and short the low sentiment, in b.p.. *, **, and *** denote significance level of 90%, 95%, and 99%, respectively. Confidence bands are constructed by performing bootstrap simulation repeatedly for 2,000 times, using sample size equal to the decile sample size of each stock.

		(1)	(2)		(3)	(4)		(5)	(6)	(7)		(8)	(9)		(10)
		$CSent_S$	$CAR0_S$	p	H-L	$CAR1_S$	p	H-L	$CSent_N$	$CAR0_N$	p	H-L	$CAR1_N$	p	H-L
AA.N	H10%	7.62	-3.88		-22.03	-6.52		19.88	11.84	-7.74		-0.54	-19.26	*	-15.38
	L10%	-6.61	18.15			-26.40	**		-19.22	-7.20			-3.88		
AAPL.OQ	H10%	45.91	32.08	***	71.49	-3.40		-8.31	63.06	33.86	***	57.56	-9.31	***	-12.80
	L10%	-37.08	-39.42			4.92			-45.33	-23.70			3.49		
AXP.N	H10%	4.89	6.82		24.35	0.68		2.91	11.67	12.08	*	21.50	2.66		2.47
	L10%	-3.44	-17.53	***		-2.23			-7.71	-9.42	*		0.19		
BA.N	H10%	9.88	11.33		8.63	9.54		7.00	27.00	16.85	**	20.16	7.26		8.43
	L10%	-15.72	2.70			2.54			-34.65	-3.31			-1.17		
BAC.N	H10%	10.36	17.39		15.88	5.47		10.82	17.22	12.40		8.24	3.85		13.10
	L10%	-36.17	1.51			-5.35			-34.44	4.17			-9.24	*	
CAT.N	H10%	6.49	8.33		28.06	2.11		15.29	13.96	11.62	*	25.37	5.83	*	12.55
	L10%	-7.34	-19.73	**		-13.19	*		-20.11	-13.75			-6.72		
CSCO.OQ	H10%	9.86	26.68		39.71	4.02		1.68	24.03	14.41		41.49	-2.87		-0.99
	L10%	-32.26	-13.03	**		2.35			-11.57	-27.08	***		-1.88		
CVX.N	H10%	6.19	-15.27		-1.01	-9.10		0.47	11.96	2.27		8.66	-1.33		6.31
	L10%	-6.15	-14.26			-9.58			-23.97	-6.39			-7.64		
DD.N	H10%	2.49	4.11		4.50	1.83		2.56	2.08	0.38		-4.48	-0.13		-3.29
	L10%	-2.92	-0.39			-0.73			-2.20	4.86			3.16		
DIS.N	H10%	7.52	8.00		12.32	6.60	*	6.46	7.83	-6.89		-5.44	-1.48		0.04
	L10%	-3.35	-4.32			0.15			-6.32	-1.45			-1.52		
GE.N	H10%	8.08	-5.25		-9.48	-0.99		-5.95	22.78	12.10	**	20.84	-0.87		0.86
	L10%	-37.11	4.23			4.96	*		-12.58	-8.74			-1.72		
GS.N	H10%	8.98	9.91		18.45	11.45		20.76	20.51	-5.02		-10.42	4.93		4.11
	L10%	-22.77	-8.54			-9.31	***		-45.40	5.40			0.82		
HD.N	H10%	8.83	19.36	**	13.61	4.90		6.68	19.87	17.33	**	14.08	2.54		0.71
	L10%	-4.21	5.75			-1.77			-7.93	3.26			1.83		
HPQ.N	H10%	9.20	20.01	**	90.34	11.14		18.41	17.54	8.15		69.57	6.12		4.31
	L10%	-16.28	-70.33	**		-7.27			-23.56	-61.42	**		1.81		
IBM.N	H10%	13.19	18.59	***	63.40	10.36	***	11.01	31.87	22.00	***	59.91	11.16	***	10.94
	L10%	-11.25	-44.82	***		-0.65			-17.81	-37.91	***		0.21		

[continue table next page]

[continue of previous table]

		(1)	(2)		(3)	(4)		(5)	(6)	(7)		(8)	(9)		(10)
		$CSent_S$	$CAR0_S$	p	H-L	$CAR1_S$	p	H-L	$CSent_N$	$CAR0_N$	p	H-L	$CAR1_N$	p	H-L
INTC.OQ	H10%	18.15	2.32		18.29	7.63		2.97	30.15	0.24		7.80	13.43	***	7.30
	L10%	-10.69	-15.97	**		4.66			-18.10	-7.57			6.13		
JNJ.N	H10%	7.26	3.15	*	11.69	-0.66		7.99	12.20	4.17	**	15.44	4.59	***	10.57
	L10%	-4.14	-8.54			-8.66	*		-10.04	-11.27	*		-5.98		
JPM.N	H10%	8.07	2.96		-3.36	2.38		-8.22	19.37	0.60		10.29	2.27		-1.64
	L10%	-19.34	6.32			10.60			-39.91	-9.68	*		3.92		
KO.N	H10%	7.32	3.26		0.77	0.28		-2.53	16.61	-5.55		6.03	-3.42		6.82
	L10%	-6.93	2.49			2.81			-12.53	-11.59			-10.24		
MCD.N	H10%	6.69	1.09		5.03	-3.43		-4.61	12.62	-3.55		-3.33	-5.29	**	-6.02
	L10%	-10.30	-3.94			1.18			-19.46	-0.23			0.72		
MMM.N	H10%	5.10	-1.92		-14.37	4.66		-2.56	11.16	-1.70		-1.70	5.52		6.45
	L10%	-2.93	12.45	***		7.21			-7.94	0.00			-0.93		
MRK.N	H10%	6.99	17.14	***	33.16	0.83		11.42	10.70	12.10	*	29.99	1.37		7.74
	L10%	-5.12	-16.02	***		-10.59	**		-10.82	-17.89	***		-6.37		
MSFT.OQ	H10%	27.42	-1.02		12.73	-7.04	**	-4.38	50.42	10.09		27.05	-2.87		1.09
	L10%	-17.56	-13.75	**		-2.66			-13.57	-16.96	***		-3.96		
NKE.N	H10%	10.28	35.48	***	135.11	2.09		9.12	11.41	-6.24		5.72	3.10		-1.62
	L10%	-4.80	-99.63	***		-7.03	**		-8.99	-11.96			4.71		
PFE.N	H10%	7.22	5.37		5.42	3.69		-4.50	12.84	1.31		16.86	-6.23	*	3.77
	L10%	-8.05	-0.05			8.18	**		-14.81	-15.55	***		-10.00	***	
PG.N	H10%	4.94	7.07	**	11.37	6.97	*	3.08	9.57	7.68	**	13.24	1.67		5.08
	L10%	-3.24	-4.30			3.89			-10.82	-5.56			-3.40		
T.N	H10%	11.18	-0.13		6.82	-0.65		5.05	17.87	-1.23		9.42	-4.13		2.45
	L10%	-9.85	-6.95			-5.70			-17.35	-10.65			-6.58		
TRV.N	H10%	1.48	2.23		-6.12	1.20		-9.15	2.30	1.51		-4.51	-8.49	**	-10.19
	L10%	-1.44	8.35	*		10.35	***		-1.53	6.02			1.70		
UNH.N	H10%	4.36	7.44		4.76	3.73		8.91	9.74	8.99		5.33	-3.02		-6.42
	L10%	-2.13	2.69			-5.18			-5.90	3.66			3.40		
UTX.N	H10%	3.85	0.24		4.78	0.20		-2.35	7.83	-2.71		-7.74	-5.88		-12.85
	L10%	-3.67	-4.54			2.55			-9.22	5.03			6.97		
V.N	H10%	6.91	16.58	**	92.72	4.39		5.93	8.05	21.71	***	15.95	1.69		-0.71
	L10%	-2.69	-76.14			-1.54			-4.63	5.76			2.39		
VZ.N	H10%	10.87	-10.46		2.53	-7.61		-1.15	19.26	-7.32		9.98	-10.07		-3.86
	L10%	-8.83	-12.99			-6.47			-17.47	-17.31			-6.21		
WMT.N	H10%	11.62	-4.46		-2.99	-4.73		-3.66	20.38	-10.88	*	-9.41	-7.93	**	-4.76
	L10%	-15.02	-1.47			-1.07			-27.74	-1.47			-3.16		
XOM.N	H10%	7.77	-7.81		3.02	0.82		11.29	16.46	-7.82		2.01	-0.42		0.96
	L10%	-10.90	-10.83			-10.48	***		-30.43	-9.83			-1.38		
Profit					**21.26**			**3.41**				**14.41**			**1.54**
Corr				**0.3926**			**0.1268**				**0.5684**			**0.1386**	

4.2 After-Hour Media Sentiment Patterns

It is well acknowledged that reports and postings in news and social media will comment on the daytime trading activity after market closed. To examining how the overnight sentiment in news and social media react to the daily returns, we perform analysis similar to the previous subsection by switching the sorting and resulting variables, and swapping the "event" windows. Again, we demonstrate the methodology in Figure 2 using IBM data. The top panel in Figure 2 shows the 10 equal-sized decile cumulative excess returns ($CAR = \sum_{t_0=9:30}^{t_1=16:00}[R_{IBM,t-1} - R_{m,t-1}]$) in trading day $t-1$ (the left-hand side), and the cumulative sentiment in **social** media from market closing to the next day's opening ($CSent_S = \sum_{t_2=16:01,(t-1)}^{t_3=9:29,t}[Sent_S]$, right-hand side of the Figure) conditional on the decile sorted CARs. Similar to the visualisation in the previous subsection, the blue thick curve indicates the highest 10% CAR and the thick red curve represents the lowest 10% CAR. We use the same colour to map into the cumulative excess returns and the corresponding cumulative sentiment. Confidence bands are constructed by performing bootstrap simulation repeating 2,000 times. The three layers of grey shaded area indicate 90%, 95%, and 99% respectively. The bottom panel of Figure 2 demonstrates the equal-sized decile sorting CARs and the corresponding **news** media cumulative sentiment after the market closed.

A comparison between the top and bottom panels of Figure 2 shows that there are both similarities and differences between social media sentiment and news media sentiment patterns in the "after-hour" period. Firstly, both panels show that the highest and lowest 10% CARs are positively associated with the after-hour cumulative sentiment. The positive and negative sentiment induced by good or bad performance are both statistically significant at 99% level, for both social media and news media. Secondly, the cumulative sentiment series builds up faster before the mid-night, with higher slopes than the slopes after the mid-night. Thirdly, the magnitudes of cumulative sentiment for social and news media reveals that the two kinds of media display different characteristics. Average cumulative social media sentiment of IBM (end point of the dashed black curve in the right-hand side of the top panel) is 1.672, whereas the average cumulative news media sentiment of IBM (end point of the dashed black curve in the right-hand side of the bottom panel) is 7.812. While the slope of $CSent_S$ after mid-night gradually flattens until the next day's opening, the most positive $CSent_N$ continues trending upward even after the mid-night with the most negative $CSent_N$ steadily becomes smooth. This result is consistent with result identified in our prior study that the tone of news media is relatively more positive than emotions in the social media.

We continue our study by performing the same analysis for all our sample DJIA stocks and summarise our results in Table 2. We reveal three main results based on Table 2. First, sentiment after market closed in social and news media are strongly positively related to the performance during trading hours. The top and bottom 10% good and bad days as measured as top and bottom decile cumulative daily excess returns are correlated with after-hour social media sentiment with coefficient equals to 0.4949, and correlated with after-hour news media sentiment with coefficient of 0.7088. And most of the impact are statistically significant at 99% level. Next, a comparison between Table 1 and Table 2 indicates that the impact of daily excess return on the after-hour sentiment is significantly stronger than the impact of overnight sentiment on the opening return. Not only do the correlation coefficients are higher, the p-values are also more significant in in Table 2 than they are in Table 2.

Last but not least, after-hour sentiment in social and news media are positively skewed. 18 out of the 34 sampling stocks contain higher positive effect than negative effect on social media (In Column (2) of Table 2, there are 18 stocks that have higher absolute value of $CSent_S$ associated with the highest 10% CAR than the absolute value of $CSent_S$ corresponding to the lowest 10% CAR). Similarly, 19 out of the 34 sampling stocks contain higher positive effect than negative effect on news media (In Column (3) of Table 2, there are 19 stocks that have higher absolute value of $CSent_N$ associated with the highest 10% CAR than the absolute value of $CSent_N$ corresponding to the lowest 10% CAR).

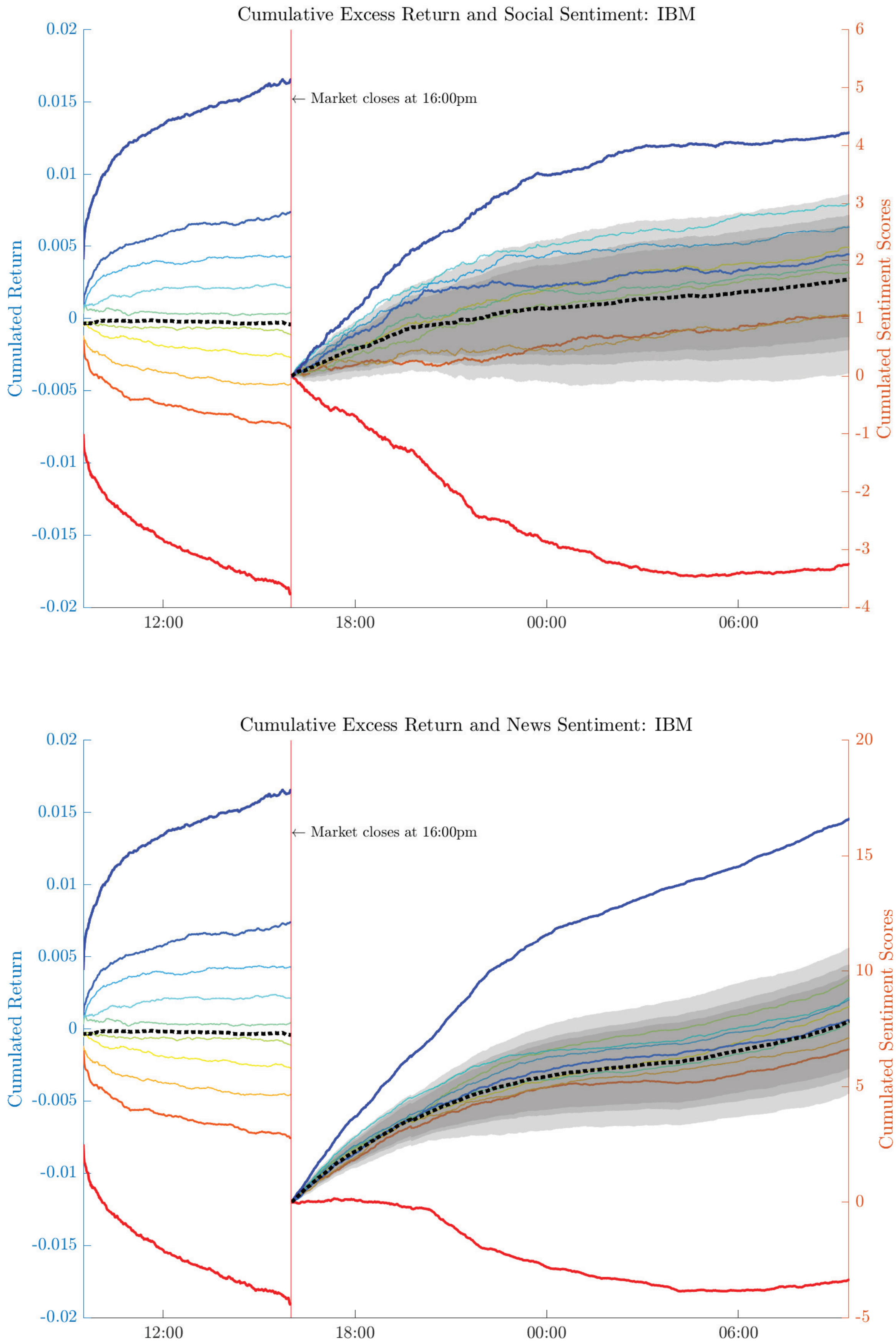


Figure 2: Example: IBM cumulative returns and next day social vs news media sentiment This plots illustrate the equally sorted cumulative returns during trading hours in the previous day (from 9:30am to 4:00pm on day *t-1*), and the associated cumulative social media sentiment (top panel) and news sentiment(bottom panel) from previous day closing (4:01pm in day *t-1*) to market open (9:29am in day *t*). Sample period: 2011/01/01-2017/11/30, and the data is at 1-minute frequency. Left-hand side curves represent cumulative returns, and the corresponding colored curves at the right-hand side are the cumulative sentiment after trading and overnight. The grey shaded confidence bands at 90%, 95% and 99% significance levels are constructed by performing bootstrap simulation using sample size of 174 (equal number of original sample size in each decile) with replacement and repeat 2,000 times.

Table 2: Extreme Cumulative Excess Return and Overnight Sentiment. This table reports the top and bottom decile cumulative excess returns of each sample stock (Column (1)), and their corresponding sorted cumulative sentiment from market close to the next trading day open in social media (Column (2)) and news media (Column (3)). Column (2) $\mathbf{CSent_S}$ and Column (3) $\mathbf{CSent_N}$ are accumulated from 16:01pm to the 9:29am. Columns (1) is the cumulative excess returns aggregated from 9:30am to 16:00pm as measured in basis points (b.p.), using the 1 minute mid-price log returns for each sample stock and subtracting mid-price log return of the DJIA index. L10% and H10% represent the average value of the lowest decile and highest decile of cumulative excess return of each stock. The correlation coefficient between extreme cumulative excess returns and the overnight social media sentiment (correlation between columns (1) and (2)), is 0.4949. The correlation coefficient between extreme cumulative excess returns and the overnight news media sentiment (correlation between columns (1) and (3)), is 0.7086. *, **, and *** denote significance level of 90%, 95%, and 99%, respectively. Confidence bands are constructed by performing bootstrap simulation repeatedly for 2,000 times, using sample size equal to the decile sample size of each stock.

		(1)	(2)		(3)				(1)	(2)		(3)	
		CAR	$CSent_S$	p	$CSent_N$	p			CAR	$CSent_S$	p	$CSent_N$	p
AA.N	L10%	-403.62	-1.99	***	-8.40	***	JPM.N	L10%	-251.12	-11.18	***	-27.77	***
	H10%	441.91	2.08	***	2.94	***		H10%	251.35	-2.83		0.37	***
AAPL.OQ	L10%	-366.01	-17.65	***	-16.96	***	KO.N	L10%	-196.07	-0.07	*	-3.04	***
	H10%	249.93	21.05	***	27.82	***		H10%	149.80	1.25	***	4.87	***
AXP.N	L10%	-212.36	-1.10	***	-3.03	***	MCD.N	L10%	-152.07	-3.21	***	-7.11	***
	H10%	205.15	1.35	***	4.42	***		H10%	151.69	1.80	***	3.63	***
BA.N	L10%	-214.21	-6.80	***	-8.36	***	MMM.N	L10%	-168.86	-0.10	***	-3.00	***
	H10%	217.93	0.25	***	10.22	***		H10%	160.73	1.25	***	5.36	***
BAC.N	L10%	-347.99	-27.04	***	-25.02	***	MRK.N	L10%	-194.82	-0.49	***	-3.81	***
	H10%	354.56	-4.18	***	2.54	***		H10%	187.70	2.63	***	4.45	***
CAT.N	L10%	-267.72	-3.08	***	-13.40	***	MSFT.OQ	L10%	-203.48	-0.27	***	10.65	***
	H10%	262.95	2.72	***	8.01	***		H10%	258.39	6.08	***	22.17	***
CSCO.OQ	L10%	-309.41	-9.71	***	-0.13	***	NKE.N	L10%	-311.67	-0.14	***	-3.01	***
	H10%	492.51	-4.36		10.64	***		H10%	236.02	3.80	***	5.41	***
CVX.N	L10%	-214.71	-2.36	***	-16.93	***	PFE.N	L10%	-179.75	-2.02	***	-5.15	***
	H10%	203.37	0.44	***	4.14	***		H10%	185.98	1.62	***	4.76	***
DD.N	L10%	-223.59	-0.91	***	-0.73	***	PG.N	L10%	-143.66	-0.27	***	-3.44	***
	H10%	213.97	-0.16		0.39	***		H10%	143.38	1.34	***	3.79	***
DIS.N	L10%	-201.23	0.37		0.53		T.N	L10%	-182.27	-1.42	***	-5.24	***
	H10%	196.86	1.95	***	0.41			H10%	154.20	1.40	**	4.67	***
GE.N	L10%	-211.29	-20.60	***	-3.13	***	TRV.N	L10%	-172.70	-0.27	***	-0.43	***
	H10%	202.32	-13.52	***	11.44	***		H10%	163.69	0.23	***	0.54	***
GS.N	L10%	-256.17	-13.01	***	-23.83	***	UNH.N	L10%	-219.83	-0.51	***	-2.85	***
	H10%	247.62	-4.01		-5.11			H10%	234.50	1.22	***	5.59	***
HD.N	L10%	-183.09	0.11		-0.53	***	UTX.N	L10%	-185.09	-0.43	***	-2.11	***
	H10%	199.41	3.53	***	9.74	***		H10%	184.76	0.72	**	2.08	***
HPQ.N	L10%	-391.53	-8.13	***	-9.71	***	V.N	L10%	-275.04	-0.27	***	-0.29	**
	H10%	317.17	-1.12		2.15	***		H10%	225.45	1.61	***	1.25	**
IBM.N	L10%	-190.66	-3.24	***	-3.35	***	VZ.N	L10%	-176.73	-0.17	***	-4.28	***
	H10%	165.65	4.22	***	16.61	***		H10%	172.93	3.41	***	7.30	***
INTC.OQ	L10%	-223.22	-2.20	***	-2.55	***	WMT.N	L10%	-177.31	-4.83	***	-12.36	***
	H10%	215.80	6.54	***	13.87	***		H10%	173.91	1.98	***	5.43	***
JNJ.N	L10%	-131.57	0.07	***	-1.93	***	XOM.N	L10%	-179.99	-4.79	***	-15.57	***
	H10%	136.51	2.44	***	5.96	***		H10%	176.50	0.15	***	4.45	***
							Corr			**0.4949**		**0.7086**	

5 Robustness Checks and Discussion

The above findings bring about two interesting insights and point out the necessity to perform robustness tests. The first insight is that our overnight sentiment measure contains two components: (1) previous day's trading performance and (2) the overnight news effects. The former component represents a positive autocorrelation of day returns, which is strongly demonstrated in the after-hour sentiment pattern part with results presented in Table 2, and the latter component indicates the "true" sentiment induced by the heterogeneity of investors, whose preferences and attention to information, as well as the associated trading behaviours based on these preferences differ, as demonstrated in Section 1. Another takeaway insights from the results with and without first minute is that, it differs among different types of stocks in terms of the speed at which stock prices fully incorporate the overnight sentiment. It would be interesting and promising to check whether such impact of overnight sentiment on returns happened faster for stocks that contain more sentiment observations than those stocks with fewer observations. An similar idea is come up with by Sul et al. (2016).

While the second insights worth further exploration which is out of the scope of this paper, we continue checking the robustness of our results in more detail following the logic of the first insight. To be more specific, we control for super good or bad performances of the previous trading day, as defined by the highest and lowest 10% cumulative excess returns CAR of each sampling stock, and repeat the opening return pattern analysis procedure - a double-sorting process that will effectively stripe away the return autocorrelation effect. Results of this robustness check is reported in Table A.7 of the Appendix in page 34. H and L under the column $CAR(t-1)$ shows the top and bottom decile daily cumulative excess returns of each stock over the sample period, other notations (H10%, L10%, $CSent_S$, $CSent_N$) have the same meaning as the previous section. To prevent from overcomplication, we only report CAR as aggregating from 9:31am to 16:00pm, and do not vary across different windows, as the daily CAR without the first minute is the most conservative results through our previous analysis.

We find that the correlation between top/bottom decile overnight investor sentiment and next day cumulative excess returns are still positive, though at lower degree, after controlling for the return autocorrelations. Correlation coefficient between social media sentiment and the corresponding CAR is equal to 0.2844, and the coefficient between news media sentiment and the associated CAR is 0.0905. Average daily cumulative excess return across all sample stocks when taking social media sentiment as a signal amounts to 15.06 b.p., much higher than that when using news media as a signal, which equals to 1.46 b.p.. Technology stocks like Apple, Microsoft, and Intel, as well as other "salient" stocks such as General Electrics and Goldman Sachs all demonstrate that the negative sentiment induces higher mispricing that positive sentiment does, consistent with Stambaugh et al. (2012).

6 Conclusion

This study investigates the descriptive property of overnight investor sentiment based on social and news media content, as well as its predictability on DJIA constituents stocks' opening returns. Using high-frequency (1-minute) investor sentiment metrics that extract and score scraped social and news media texts via a proprietary algorithm from Thomson Reuters MarketPsych Indices (TRMI), we study whether overnight investor sentiment predict opening return the next trading day. We find that the cumulative excess returns of DJIA constituent stocks are significantly positively related to the top and bottom decile overnight sentiment from social and news media. Varying various time windows - the first half hour, the first hour, and morning session of the next trading day, we find consistent results that the cumulative excess returns are positively correlated with overnight **social** media sentiment around a coefficient of 0.4, irrespective of the event window. Such correlation coefficient between **news** media sentiment and opening return is higher, and ranges from 0.54 to 0.56. Getting rid of the first trading minute and continuing examining such associations, we find that these associations become weaker. Specifically, the correlation coefficients between extreme overnight **social** media sentiment and cumulative excess returns all diminish to approximately 0.1 for the first half hour, the first hour, and morning session. Such correlation coefficients for sentiment expressed in **news** media also decreased to approximately 0.1. These findings suggest that the overnight return (open-to-close) is an appropriate proxy for individual stock sentiment, and is consistent with Aboody et al. (2018). Moreover, the quick diminishing effect also implies that overnight sentiment is swiftly impounded into stock prices in the first minute of trading. These results are robust when we controlled for the return autocorrelation influences. Using overnight social media sentiment as a signal, the sample average daily cumulative excess return is at least 15.06 b.p., which is much higher than the taking news based sentiment as a signal.

We proceed this study by conducting equal-sized decile sorting of the cumulative excess returns and scrutinising the corresponding after-hour investor sentiment in social and news media. We find that the extremely good (bad) trading days, measured as the highest and lowest 10% daily cumulative excess returns, are highly positively connected with positive (negative) emotions in social and media after the market closed. The correlation coefficient between extremely good (bad) days and the associated after-hour social media sentiment is approximately 0.50, while this correlation coefficient for the news media sentiment is at 0.71, which is significantly higher than for the social media sentiment. Besides, this "reverse" effect from market to media sentiment is statistically more prominent than the effect from overnight sentiment to market, based on our bootstrap simulation results. We also observe that for most sampling stocks investigated, the build-up of sentiment during non-trading hours is faster between 18:00pm and the mid-night than other periods, whereas the curve of cumulative sentiment continues increasing at a slower rate after the mid-night, or even flattens out. This finding is consistent with the daily routines and regular posting behaviours of most individual investors, which provides support that the TRMI data is effective in capturing social media and news media activities over-the-clock.

We underscore the differences between social media sentiment and news based sentiment, and document that both their predictive and descriptive properties differs from each other. Combining our previous findings, we provide implication that overnight sentiment expressed in social and news media contains two important components that drive opening return the next trading day: a component from previous day's trading performance and the behavioural heterogeneity among different investors. Overall, using individual stock specific sentiment measures, this paper contributes to the literature that investigates the overnight investor sentiment and intraday return patterns. It also provides new evidence that the asymmetry between positive and negative sentiment effects mainly comes from the short-leg, and brings about new insights to how information is incorporated into prices in response to the increasing prominence of social media as information dissemination channel.

References

Aboody, D., Even-Tov, O., Lehavy, R., and Trueman, B. (2018). Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis*, 53(2):485–505.

Akhtar, S., Faff, R., Oliver, B., and Subrahmanyam, A. (2012). Stock salience and the asymmetric market effect of consumer sentiment news. *Journal of Banking & Finance*, 36(12):3289–3301.

Bagnoli, M., Clement, M. B., and Watts, S. G. (2005). Around-the-clock media coverage and the timing of earnings announcements. *Working Paper*.

Barber, B. M. and Odean, T. (2007). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, 21(2):785–818.

Barber, B. M., Odean, T., and Zhu, N. (2008). Do retail trades move markets? *The Review of Financial Studies*, 22(1):151–186.

Behrendt, S. and Schmidt, A. (2018). The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility. *Journal of Banking & Finance*, 96:355–367.

Berkman, H., Koch, P. D., Tuttle, L., and Zhang, Y. J. (2012). Paying attention: overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis*, 47(4):715–741.

Birru, J. (2018). Day of the week and the cross-section of returns. *Journal of Financial Economics*, 130(1):182–214.

Bradley, D., Clarke, J., Lee, S., and Ornthanalai, C. (2014). Are analysts' recommendations informative? intraday evidence on the impact of time stamp delays. *The Journal of Finance*, 69(2):645–673.

Branch, B. and Ma, A. (2012). Overnight return, the Invisible Hand Behind Intraday Returns? *Journal of Applied Finance*, 22(2):90–100.

Chen, H., Noronha, G., and Singal, V. (2004). The price response to s&p 500 index additions and deletions: Evidence of asymmetry and a new explanation. *The Journal of Finance*, 59(4):1901–1930.

Cooper, M. J., Cliff, M. T., and Gulen, H. (2008). Return differences between trading and non-trading hours: Like night and day. *SSRN Electronic Journal*.

Engelberg, J., Sasseville, C., and Williams, J. (2012). Market madness? the case of mad money. *Management Science*, 58(2):351–364.

Gao, L., Han, Y., Li, S. Z., and Zhou, G. (2018). Market intraday momentum. *Journal of Financial Economics*, 129(2):394–414.

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

Harris, L. (1986). A transaction data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics*, 16(1):99–117.

Hong, H. and Wang, J. (2000). Trading and returns under periodic market closures. *The Journal of Finance*, 55(1):297–354.

Jain, P. C. and Joh, G.-H. (1988). The dependence between hourly prices and trading volume. *The Journal of Financial and Quantitative Analysis*, 23(3):269.

Jiang, C. X., Likitapiwat, T., and McInish, T. H. (2012). Information content of earnings announcements: Evidence from after-hours trading. *Journal of Financial and Quantitative Analysis*, 47(6):1303–1330.

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

Kelley, E. K. and Tetlock, P. C. (2013). Why do investors trade? *Working Paper*.

Kelly, M. A. and Clark, S. P. (2011). Returns in trading versus non-trading hours: The difference is day and night. *Journal of Asset Management*, 12(2):132–145.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Lou, D., Polk, C., and Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics*.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39.

Michaely, R., Rubin, A., and Vedrashko, A. (2013). Corporate governance and the timing of earnings announcements. *Review of Finance*, 18(6):2003–2044.

Palomino, F., Renneboog, L., and Zhang, C. (2009). Information salience, investor sentiment, and stock returns: The case of british soccer betting. *Journal of Corporate Finance*, 15(3):368–387.

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

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2014). The long of it: Odds that investor sentiment spuriously predicts anomaly returns. *Journal of Financial Economics*, 114(3):613–619.

Sul, H. K., Dennis, A. R., and Yuan, L. I. (2016). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48(3):454–488.

Sun, L., Najand, M., and Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance*, 73:147–164.

Thaler, R. (1987). Anomalies: Seasonal movements in security prices II: Weekend, holiday, turn of the month, and intraday effects. *Journal of Economic Perspectives*, 1(2):169–177.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232.

A Appendix

A.1 List of acronyms

Table A.1: List of acronyms

Acronym	Description
BW	Baker and Wurgler
DJIA	Dow Jones Industry Average
ECN	Electronic Communication Network
ETF	Exchange-traded Funds
GFC	Global Financial Crisis (from 2008 to 2010)
ICS	Index of Consumer Sentiment
LSE	London Stock Exchage
NYSE	New York Stock Exchange
SIRCA	Securities Industry Research Centre of Asia-Pacific
S&P	Standard and Poor
TRMI	Thomson Reuters MarketPsych Indices
TRNA	Thomson Reuters News Analytics
TRTH	Thomson Reuters Tick History
US	The United States
VWAP	Volume weighted average price

A.2 Data sources and variable names

Table A.2: List of data sources and variable names. This table summarises variables used and their definition.

Code/Symbol	Description
DataScope	Thomson Reuters DATASCOPE
TRMI	Thomson Reuters MarketPsych Indices
MPTRXUS30	TRMI company group code (DJIA respective sentiment)
$P^m_{t,j}$	mid-prices of stock on date t at time j
$P^l_{t,j}$	last-prices of stock on date t at time j
$R_{t,j}$	mid-price return of stock on date t at time j
$R^l_{t,j}$	last-price return of stock on date t at time j
R_m	mid-price market return as calculated from the DJIA index
R^l_m	last-price market return as calculated from the DJIA index
$A_{t,j}$	ask prices of stock on date t at time j
$B_{t,j}$	bid prices of stock on date t at time j
K	media type indicator: $K = S$ social media, $K = N$ news media
AR	abnormal return
CAR	cumulative abnormal return
$ACAR$	average cumulative abnormal return
$CSent_S$	cumulative social media sentiment for a specific time frame
$CSent_N$	cumulative news media sentiment for a specific time frame
CR_{noon}	cumulative return from market open (9:30am) to noon (12:00pm)
CR_{1H}	cumulative return for the first trading hour (from 9:30am to 10:30am)

A.3 Social and News Activities for Sample Stocks

Table A.3: TRMI SAMPLE NUMBER OF OBSERVATIONS. This table lists TRMI sample size for social and news media activity volume and the emotions for each individual stock and the overall Dow Jones company groups. These observations are irregular and before being dealt with by the retime procedure. $Buzz_S$ and $Buzz_N$ are measures that capture total volumes of social media or news activities. $Sent_S$ and $Sent_N$ are the net positive and negative emotional scores for each entities from social media and news, respectively. Sample period: 2011/01/01-2017/11/30 at 1-minute frequency.

RIC	$Buzz_S$	$Buzz_N$	$Sent_S$	$Sent_N$	**RIC**	$Buzz_S$	$Buzz_N$	$Sent_S$	$Sent_N$
AA.N	77,541	54,850	64,063	50,369	KFT.OQ	9,103	26,750	6,726	22,658
AAPL.OQ	1,476,678	983,446	1,310,025	910,719	KO.N	85,066	141,893	69,217	126,629
AXP.N	28,943	57,471	22,970	49,300	MCD.N	101,715	145,284	83,752	130,989
BA.N	196,935	331,032	168,487	292,763	MMM.N	38,514	60,766	30,326	52,848
BAC.N	463,226	227,393	400,181	195,850	MRK.N	69,885	73,191	56,075	63,800
CAT.N	68,265	61,293	57,194	55,463	MSFT.OQ	429,844	564,742	361,855	507,409
CSCO.OQ	340,545	149,162	300,459	132,024	NKE.N	65,722	64,843	52,647	57,582
CVX.N	53,402	112,879	43,411	97,178	PFE.N	113,727	103,159	94,373	89,748
DD.N	23,156	7,857	19,965	6,592	PG.N	39,585	64,748	33,208	58,429
DIS.N	41,484	43,998	33,652	38,117	T.N	190,843	178,099	159,040	151,011
GE.N	445,679	202,292	390,059	173,480	TRV.N	6,290	5,761	4,520	5,107
GS.N	337,229	368,779	291,235	320,741	UNH.N	16,843	30,028	13,058	25,630
HD.N	51,712	60,676	41,674	54,084	UTX.N	20,132	35,870	15,836	30,595
HPQ.N	169,747	192,543	146,304	170,659	V.N	35,036	21,529	27,532	19,075
IBM.N	138,948	223,869	112,768	198,993	VZ.N	145,293	183,045	116,153	154,311
INTC.OQ	263,700	232,588	224,186	204,624	WMT.N	250,033	237,907	212,873	212,538
JNJ.N	71,096	79,074	57,250	68,966	XOM.N	131,756	172,538	109,729	151,723
JPM.N	72,096	359,119	192,823	311,167	**.DJI**	**2,753,603**	**2,536,911**	**2,593,029**	**2,449,177**

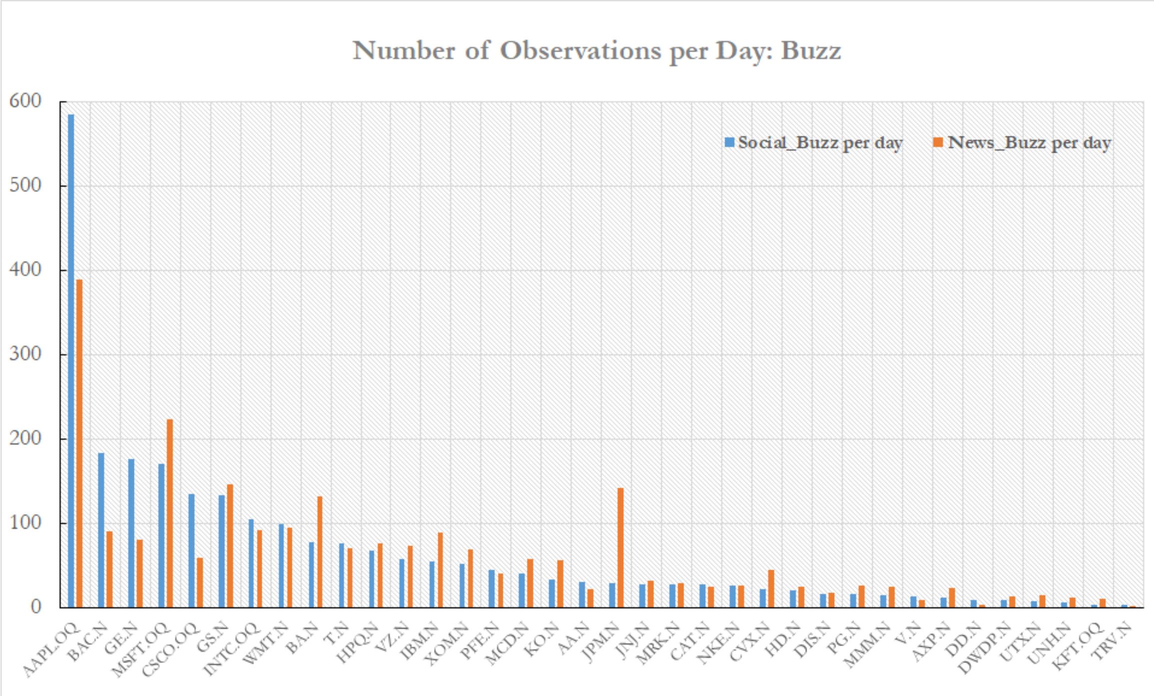


Figure A.1: Sample Stock Buzz Observations Per Day. This figure display number of observations on the volume of postings (*Buzz*) in media of each sample stock, before we refill empty values and reshaping the time scale. For each stock, the sample period is from 2011/01/01 to 2017/11/30. Blue bars indicate social media activities, and orange bars represent news media activities.

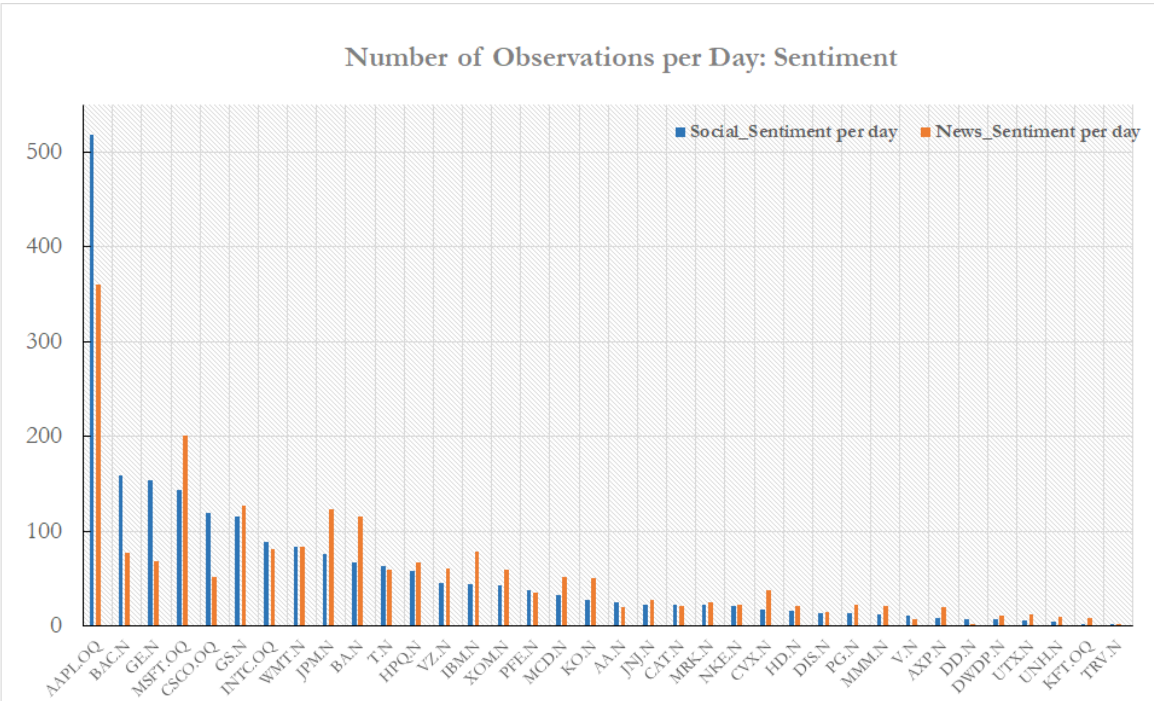


Figure A.2: Sample Stock Sentiment Observations Per Day. This figure display number of observations on the net emotional scores (*Sentiment*) in media of each sample stock, before we refill empty values and reshaping the time scale. For each stock, the sample period is from 2011/01/01 to 2017/11/30. Blue bars indicate social media sentiment, and orange bars represent news media sentiment.

A.4 Summary of Events Sample Distribution

Table A.4: Overnight Sentiment Events Sample Distribution. This table shows the distribution of overnight sentiment "event" for each sample stock. Sentiment series are aggregated from 00:01am to 09:29am each day. Days of over 95% zero-returns are excluded, and the number of such days are summarized in the column "Removed". The column "Remain" is the total number of events for each sample stock we obtain after the data pre-processing procedure. Panel A shows the distribution of aggregated social media sentiment events, and Panel B is a description of the aggregated news media sentiment events.

Panel A : Summary and Distribution of Sample Events - Social Media

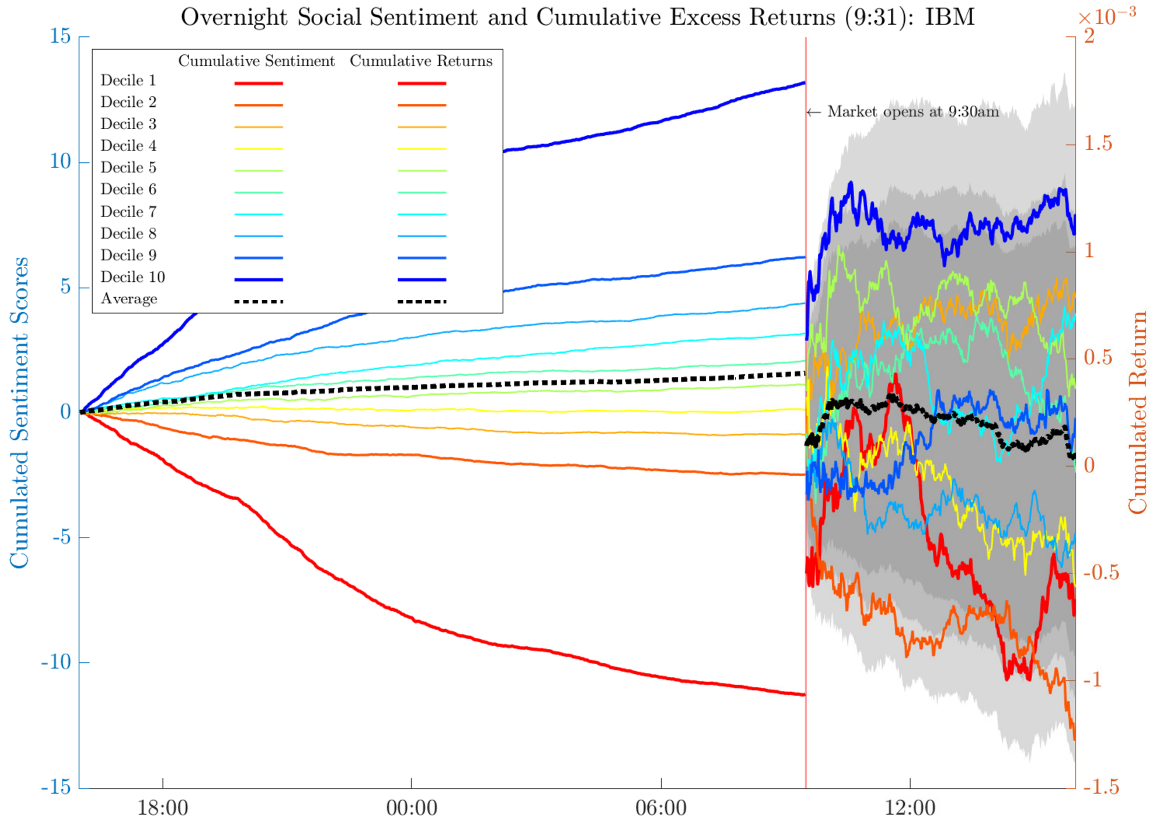
RIC	Removed	Events	Negative				Neutral		Positive			
			10th	20th	30th	40th	50th	60th	70th	80th	90th	100th
AA.N	905	1,622	162	162	163	162	227	97	166	159	162	162
AAPL.OQ	786	1,741	174	174	174	174	174	175	174	174	174	174
AXP.N	786	1,741	174	174	622	0	0	35	178	170	175	173
BA.N	786	1,741	174	174	174	174	174	175	174	174	174	174
BAC.N	823	1,705	170	171	170	171	170	171	171	170	171	170
CAT.N	785	1,741	174	174	174	178	282	63	197	151	174	174
CSCO.OQ	786	1,741	174	174	174	174	174	175	174	174	174	174
CVX.N	786	1,741	174	174	174	174	236	119	168	174	174	174
DD.N	849	1,678	168	168	167	168	513	0	0	159	167	168
DIS.N	786	1,741	178	170	174	397	0	130	203	144	171	174
GE.N	791	1,736	174	173	174	173	174	174	173	174	173	174
GS.N	786	1,741	174	174	174	174	175	174	174	174	174	174
HD.N	785	1,741	174	174	175	310	38	174	174	174	174	174
HPQ.N	791	1,736	174	173	174	173	174	174	173	174	173	174
IBM.N	785	1,741	174	174	174	174	175	174	175	173	174	174
INTC.OQ	786	1,741	174	174	174	174	175	174	174	174	174	174
JNJ.N	785	1,741	174	174	174	174	189	160	174	174	174	174
JPM.N	786	1,741	174	174	174	174	175	174	174	174	174	174
KFT.OQ	2,457	68	7	7	40	0	0	0	0	0	7	7
KO.N	786	1,741	174	174	174	174	175	174	174	174	174	174
MCD.N	785	1,741	174	174	174	174	175	174	174	174	174	174
MMM.N	786	1,741	181	178	172	286	54	174	174	174	175	173
MRK.N	786	1,741	174	182	168	268	80	173	174	174	174	174
MSFT.OQ	786	1,741	174	174	174	174	175	174	174	174	174	174
NKE.N	785	1,741	174	196	152	301	48	174	174	174	174	174
PFE.N	786	1,741	174	174	174	174	175	174	174	174	174	174
PG.N	785	1,741	174	174	175	239	113	170	174	174	174	174
T.N	786	1,741	174	174	174	174	175	174	174	174	174	174
TRV.N	785	1,739	179	1281	0	0	0	0	0	0	112	167
UNH.N	784	1,741	174	1043	0	0	0	0	2	203	210	109
UTX.N	784	1,741	190	159	802	0	0	0	68	182	208	132
V.N	786	1,741	174	177	601	0	0	94	187	215	119	174
VZ.N	786	1,741	174	174	174	174	175	174	174	174	174	174
WMT.N	786	1,741	174	174	174	174	175	174	174	174	174	174
XOM.N	786	1,741	174	174	174	174	175	178	173	171	174	174
Total Events		**59,032**										

[continue table next page]

[continue of previous table]

Panel B : Summary and Distribution of Sample Events - News Media

RIC	Removed	Events	Negative				Neutral		Positive			
			10th	20th	30th	40th	50th	60th	70th	80th	90th	100th
AA.N	905	1,622	162	185	140	688	0	0	0	123	162	162
AAPL.OQ	786	1,741	174	174	174	174	174	175	174	174	174	174
AXP.N	786	1,741	174	174	354	0	178	165	174	174	174	174
BA.N	786	1,741	174	174	174	174	174	175	174	174	174	174
BAC.N	823	1,705	170	171	170	171	170	171	171	170	171	170
CAT.N	785	1,741	174	174	180	332	11	174	174	174	174	174
CSCO.OQ	786	1,741	174	174	174	174	174	175	174	174	174	174
CVX.N	786	1,741	174	174	174	174	176	173	174	174	174	174
DD.N	849	1,677	172	1,308	0	0	0	0	0	0	29	168
DIS.N	786	1,741	192	159	177	601	0	0	90	174	174	174
GE.N	791	1,736	174	173	174	173	174	174	173	174	173	174
GS.N	786	1,736	174	174	174	174	175	174	174	174	174	174
HD.N	786	1,741	174	174	176	359	0	162	175	173	174	174
HPQ.N	791	1,736	174	173	174	173	174	174	173	174	173	174
IBM.N	785	1,741	174	174	174	174	175	174	174	174	174	174
INTC.OQ	786	1,741	174	174	174	174	175	174	174	174	174	174
JNJ.N	786	1,741	174	174	174	284	65	174	174	180	168	174
JPM.N	786	1,741	174	174	174	174	175	174	174	174	174	174
KFT.OQ	2,458	68	8	6	6	21	0	0	7	6	7	7
KO.N	786	1,741	174	174	174	174	175	174	174	174	174	174
MCD.N	786	1,741	174	174	174	174	175	174	174	174	174	174
MMM.N	786	1,741	174	174	175	173	179	182	163	173	174	174
MRK.N	786	1,741	174	174	174	298	51	174	175	174	173	174
MSFT.OQ	786	1,741	174	174	174	174	175	174	174	174	174	174
NKE.N	786	1,741	174	174	175	258	90	174	177	171	174	174
PFE.N	786	1,741	174	174	174	174	226	123	174	174	174	174
PG.N	786	1,741	174	174	174	216	133	174	174	174	174	174
T.N	786	1,741	174	174	174	174	175	174	174	174	174	174
TRV.N	784	1,738	1648	0	0	0	0	0	0	0	0	90
UNH.N	786	1,741	175	1057	0	0	0	0	0	161	211	137
UTX.N	785	1,741	192	161	766	0	0	0	100	174	174	174
V.N	784	1,741	174	174	882	0	0	0	0	163	174	174
VZ.N	786	1,741	174	174	174	174	175	174	174	174	174	174
WMT.N	786	1,741	174	174	174	174	175	174	174	174	174	174
XOM.N	786	1,741	174	174	174	174	175	174	174	174	174	174
Total Events		**59,025**										



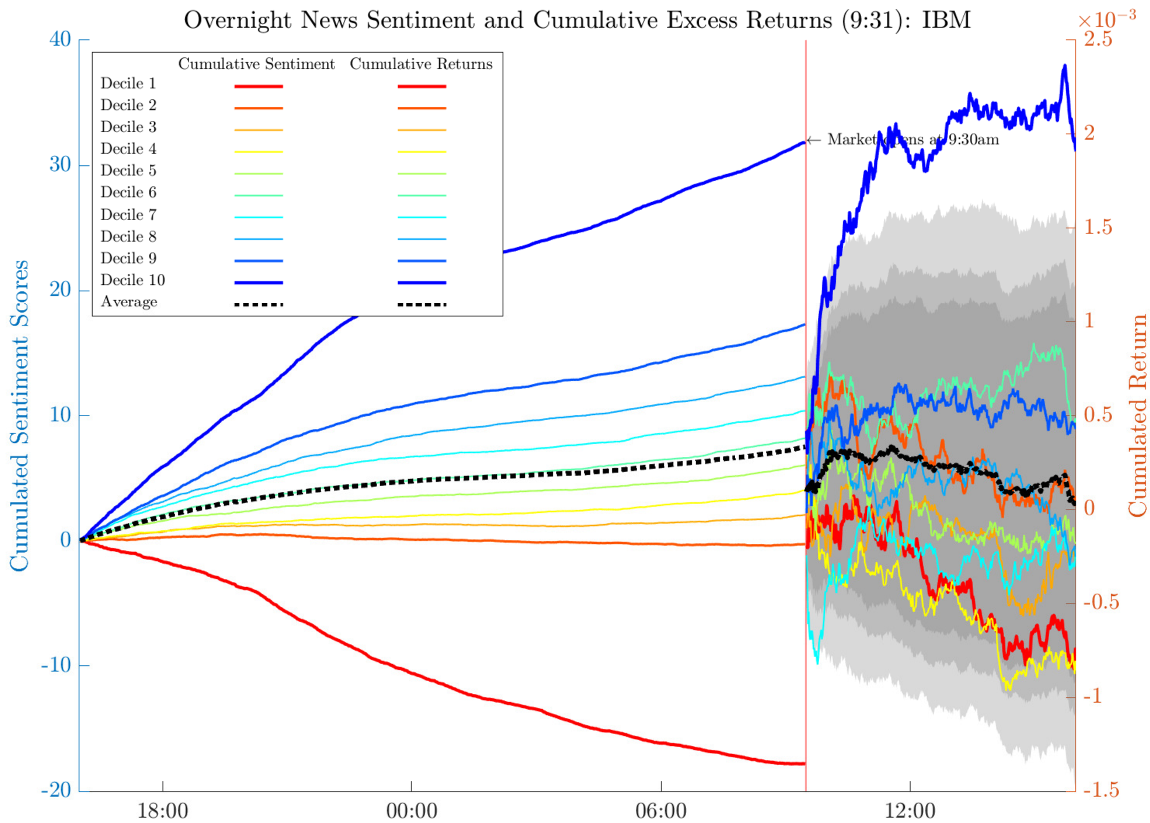


Figure A.3: Example: IBM overnight media sentiment and next day cumulative returns (9:31am) This plots illustrate the equally sorted cumulative social media sentiment (top panel) and news sentiment(bottom panel) from previous day closing (4:00pm on day *t-1*) to market open (9:29am on day *t*), and the associated next trading day cumulative returns, accumulated from 9:31am to 4:00pm on day *t*. Sample period: 2011/01/01-2017/11/30, and the data is at 1-minute frequency. Left-hand side curves represent cumulative sentiment, and the corresponding colored curves at the right-hand side are the cumulative returns next day. The grey shaded confidence bands at 90%, 95% and 99% significance levels are constructed by performing bootstrap simulation using sample size of 174 (equal number of original sample size in each decile) with replacement and repeat 2,000 times.

A.5 Different Event Windows

Table A.5: Extreme Overnight Sentiment and First Hour Cumulative Excess Return. This table reports the top and bottom decile cumulative sentiment of each sample stock from previous day market close to current day market open, and their corresponding cumulative excess returns during the first trading hour. Column (1) $\mathbf{CSent_S}$ and Column (6) $\mathbf{CSent_N}$ are cumulative sentiment of social and news media respectively, aggregating from 16:01pm to the 9:29am. Columns (2) $\mathbf{CAR0_S}$ and (7) $\mathbf{CAR0_N}$ are the cumulative excess returns aggregated from 9:30am to 10:30am as measured in basis points (b.p.), using the 1 minute mid-price log returns for each sample stock and subtracting mid-price log returns of the DJIA index. Columns (4) $\mathbf{CAR1_S}$ and (9) $\mathbf{CAR1_N}$ are the corresponding cumulative excess returns of each stock aggregated from 9:31am to 10:30am in basis point (b.p.). H10% and L10% represent the average value of the top and bottom decile cumulative sentiment. The correlation coefficient between strong social media sentiment and $CAR0_S$, including the first minute (the correlation between columns (1) and (2)), is 0.3759. The correlation coefficient between strong overnight news sentiment and $CAR0_S$, including the first minute (columns (6) and (7)), is 0.5147. Eliminating the first minute (9:30am), the correlation between strong social media sentiment and the first half hour CAR (column (1) and (4)) is only 0.0758, while the correlation between extreme news sentiment and the corresponding first half hour CAR (columns (6) and (9)) is 0.0304. Columns (3), (5), (8) and (10) "H-L" are the differences if we long highest sentiment CAR and short low sentiment CAR of each stock, expressed in columns (2), (4), (7), and (9). "Profit" indicates the average CAR differences across sample stocks when one longs the high sentiment and short the low sentiment, in b.p.. *, **, and *** denote significance level of 90%, 95%, and 99%, respectively. Confidence bands are constructed by performing bootstrap simulation repeatedly for 2,000 times, using sample size equal to the decile sample size of each stock.

		(1)	(2)		(3)	(4)		(5)	(6)	(7)		(8)	(9)		(10)
		$CSent_S$	$CAR0_S$	p	H-L	$CAR1_S$	p	H-L	$CSent_N$	$CAR0_N$	p	H-L	$CAR1_N$	p	H-L
AA.N	H10%	7.62	0.96		-20.20	-1.68	**	21.71	11.84	-9.36		-5.16	-20.88		-19.99
	L10%	-6.61	21.16			-23.39			-19.22	-4.20			-0.89		
AAPL.OQ	H10%	45.91	34.52	***	71.26	-0.95		-8.55	63.06	30.51	***	51.42	-12.66	**	-18.94
	L10%	-37.08	-36.74			7.60			-45.33	-20.91			6.28		
AXP.N	H10%	4.89	3.20		19.81	-2.93		-1.63	11.67	12.54	*	24.13	3.12		5.10
	L10%	-3.44	-16.60	***		-1.30			-7.71	-11.58	**		-1.98		
BA.N	H10%	9.88	11.54		9.52	9.74		7.89	27.00	20.26	**	23.86	10.67		12.13
	L10%	-15.72	2.02			1.86			-34.65	-3.60			-1.46	*	
BAC.N	H10%	10.36	15.98		12.53	4.07		7.48	17.22	9.56		-3.50	1.01		1.36
	L10%	-36.17	3.45			-3.41			-34.44	13.06			-0.35	*	
CAT.N	H10%	6.49	4.96		28.65	-1.27		15.88	13.96	15.10	*	34.49	9.31		21.67
	L10%	-7.34	-23.70	**		-17.15	**		-20.11	-19.39	*		-12.37		
CSCO.OQ	H10%	9.86	24.01		36.56	1.36		-1.47	24.03	13.65		48.25	-3.63		5.78
	L10%	-32.26	-12.55	**		2.83			-11.57	-34.61	***		-9.41	**	
CVX.N	H10%	6.19	-14.13		-6.32	-7.97		-4.83	11.96	3.74	*	9.26	0.13		6.91
	L10%	-6.15	-7.81			-3.13			-23.97	-5.52			-6.78	*	
DD.N	H10%	2.49	1.45		2.85	-0.83		0.91	2.08	0.62		-1.58	0.10		-0.39
	L10%	-2.92	-1.40			-1.74			-2.20	2.20			0.50	*	
DIS.N	H10%	7.52	8.97	*	12.91	7.57	*	7.05	7.83	-4.23		-1.95	1.17		3.52
	L10%	-3.35	-3.94			0.52			-6.32	-2.28			-2.35		
GE.N	H10%	8.08	-6.00		-9.76	-1.73		-6.22	22.78	13.28	***	22.49	0.32		2.51
	L10%	-37.11	3.76			4.49			-12.58	-9.21			-2.19		
GS.N	H10%	8.98	7.78		19.88	9.32		22.19	20.51	-5.84	***	-11.70	4.10	*	2.83
	L10%	-22.77	-12.10	*		-12.87			-45.40	5.85			1.27		
HD.N	H10%	8.83	23.25		19.37	8.80		12.44	19.87	15.86		6.35	1.06		-7.01
	L10%	-4.21	3.88			-3.65			-7.93	9.50			8.08		
HPQ.N	H10%	9.20	20.28	**	86.86	11.41		14.93	17.54	9.71		70.47	7.68		5.20
	L10%	-16.28	-66.58	**		-3.52			-23.56	-60.75	**		2.48		

[continue table next page]

[continue of previous table]

		(1)	(2)		(3)	(4)		(5)	(6)	(7)		(8)	(9)		(10)
		$CSent_S$	$CAR0_S$	p	H-L	$CAR1_S$	p	H-L	$CSent_N$	$CAR0_N$	p	H-L	$CAR1_N$	p	H-L
IBM.N	H10%	13.19	20.89	***	63.74	12.67	**	11.34	31.87	25.68	***	64.29	14.83	***	15.32
	L10%	-11.25	-42.84	***		1.33	*		-17.81	-38.61	***		-0.49		
INTC.OQ	H10%	18.15	3.95		24.96	9.27		9.64	30.15	-0.14		10.17	13.05	**	9.66
	L10%	-10.69	-21.01	***		-0.37			-18.10	-10.31			3.39		
JNJ.N	H10%	7.26	2.57	*	13.56	-1.24		9.86	12.20	5.39	**	15.30	5.81	***	10.43
	L10%	-4.14	-10.99			-11.11	*		-10.04	-9.91			-4.62		
JPM.N	H10%	8.07	-0.41		-10.42	-0.99	*	-15.28	19.37	-2.27		5.30	-0.60		-6.63
	L10%	-19.34	10.01			14.29			-39.91	-7.57	*		6.03		
KO.N	H10%	7.32	2.84		-3.37	-0.14		-6.67	16.61	-3.82		3.35	-1.69		4.14
	L10%	-6.93	6.21			6.53			-12.53	-7.17			-5.82		
MCD.N	H10%	6.69	6.01		8.78	1.49		-0.86	12.62	-1.74		-1.99	-3.48	*	-4.68
	L10%	-10.30	-2.77			2.35			-19.46	0.25			1.20		
MMM.N	H10%	5.10	-4.32		-15.77	2.25		-3.95	11.16	-3.23		-4.49	3.99		3.66
	L10%	-2.93	11.45	**		6.20			-7.94	1.26			0.33		
MRK.N	H10%	6.99	17.08	***	34.36	0.77		12.62	10.70	12.55	*	30.92	1.82		8.67
	L10%	-5.12	-17.28	***		-11.85	**		-10.82	-18.37	***		-6.85		
MSFT.OQ	H10%	27.42	-2.61		12.63	-8.63	*	-4.48	50.42	4.05		20.86	-8.90		-5.10
	L10%	-17.56	-15.24	**		-4.15			-13.57	-16.81	**		-3.80		
NKE.N	H10%	10.28	31.44	***	131.95	-1.95		5.95	11.41	-7.75		2.41	1.58		-4.93
	L10%	-4.80	-100.51	***		-7.90	*		-8.99	-10.16			6.52		
PFE.N	H10%	7.22	9.95		13.80	8.26		3.88	12.84	3.20		19.23	-4.33		6.15
	L10%	-8.05	-3.85			4.38			-14.81	-16.03	***		-10.48	**	
PG.N	H10%	4.94	4.70	*	5.45	4.60	*	-2.84	9.57	7.78	*	13.48	1.77		5.31
	L10%	-3.24	-0.75			7.44			-10.82	-5.69			-3.54		
T.N	H10%	11.18	-4.74		0.95	-5.25		-0.81	17.87	-2.45	*	8.21	-5.36		1.24
	L10%	-9.85	-5.69			-4.44			-17.35	-10.66			-6.60		
TRV.N	H10%	1.48	1.03		-6.85	0.00		-9.89	2.30	1.14		-2.33	-8.86	**	-8.01
	L10%	-1.44	7.88			9.89	**		-1.53	3.47			-0.85		
UNH.N	H10%	4.36	12.46		8.17	8.74		12.32	9.74	9.61		11.37	-2.40		-0.39
	L10%	-2.13	4.28			-3.58			-5.90	-1.75			-2.01		
UTX.N	H10%	3.85	3.12		9.56	3.08		2.43	7.83	-1.22		-4.06	-4.39	*	-9.17
	L10%	-3.67	-6.44			0.65			-9.22	2.85			4.78		
V.N	H10%	6.91	16.08	**	95.46	3.89		8.68	8.05	24.32	***	15.35	4.30		-1.30
	L10%	-2.69	-79.39	*		-4.79			-4.63	8.97			5.60		
VZ.N	H10%	10.87	-12.08		-1.26	-9.23		-4.93	19.26	-5.83		10.16	-8.57		-3.68
	L10%	-8.83	-10.82			-4.29			-17.47	-15.99	**		-4.89		
WMT.N	H10%	11.62	-4.03		-6.27	-4.30	*	-6.95	20.38	-9.64		-10.17	-6.69	**	-5.52
	L10%	-15.02	2.24			2.64			-27.74	0.53			-1.16		
XOM.N	H10%	7.77	-9.64		-3.36	-1.01		4.91	16.46	-8.55		-0.54	-1.16		-1.58
	L10%	-10.90	-6.27			-5.92	**		-30.43	-8.02			0.43		
Profit					**20.61**			**2.76**				**14.51**			**1.64**
Corr			**0.3759**			**0.0758**				**0.5147**			**0.0304**		

Table A.6: Extreme Overnight Sentiment and Morning Cumulative Excess Return. This table reports the top and bottom decile cumulative sentiment of each sample stock from previous day market close to current day market open, and their corresponding cumulative excess returns in the morning of next trading day. Column (1) $\mathbf{CSent_S}$ and Column (6) $\mathbf{CSent_N}$ are cumulative sentiment of social and news media respectively, aggregating from 16:01pm to the 9:29am. Columns (2) $\mathbf{CAR0_S}$ and (7) $\mathbf{CAR0_N}$ are the cumulative excess returns aggregated from 9:30am to 12:00pm as measured in basis points (b.p.), using the 1 minute mid-price log returns for each sample stock and subtracting mid-price log returns of the DJIA index. Columns (4) $\mathbf{CAR1_S}$ and (9) $\mathbf{CAR1_N}$ are the corresponding cumulative excess returns of each stock aggregated from 9:31am to 12:00pm in basis point (b.p.). H10% and L10% represent the average value of the top and bottom decile cumulative sentiment. The correlation coefficient between strong social media sentiment and $CAR0_S$, including the first minute (the correlation between columns (1) and (2)), is 0.3997. The correlation coefficient between strong overnight news sentiment and $CAR0_S$, including the first minute (columns (6) and (7)), is 0.5464. Eliminating the first minute (9:30am), the correlation between strong social media sentiment and the first half hour CAR (column (1) and (4)) is only 0.0969, while the correlation between extreme news sentiment and the corresponding first half hour CAR (columns (6) and (9)) is 0.1191. Columns (3), (5), (8) and (10) "H-L" are the differences if we long highest sentiment CAR and short low sentiment CAR of each stock, expressed in columns (2), (4), (7), and (9). "Profit" indicates the average CAR differences across sample stocks when one longs the high sentiment and short the low sentiment, in b.p.. *, **, and *** denote significance level of 90%, 95%, and 99%, respectively. Confidence bands are constructed by performing bootstrap simulation repeatedly for 2,000 times, using sample size equal to the decile sample size of each stock.

		(1)	(2)		(3)	(4)		(5)	(6)	(7)		(8)	(9)		(10)
		$CSent_S$	$CAR0_S$	p	H-L	$CAR1_S$	p	H-L	$CSent_N$	$CAR0_N$	p	H-L	$CAR1_N$	p	H-L
AA.N	H10%	7.62	-0.36		-17.36	-3.00	**	24.55	11.84	-6.91		18.40	-18.43		3.57
	L10%	-6.61	16.99			-27.55			-19.22	-25.31			-22.00		
AAPL.OQ	H10%	45.91	37.78	***	71.19	2.31		-8.62	63.06	35.06	***	55.99	-8.10	**	-14.36
	L10%	-37.08	-33.41			10.93			-45.33	-20.93			6.26		
AXP.N	H10%	4.89	0.52		13.79	-5.61	*	-7.65	11.67	13.03		22.03	3.60		3.00
	L10%	-3.44	-13.27	**		2.03	*		-7.71	-9.00			0.60		
BA.N	H10%	9.88	11.72		8.06	9.92		6.43	27.00	15.25	**	14.39	5.66		2.67
	L10%	-15.72	3.66			3.50			-34.65	0.85			2.99		
BAC.N	H10%	10.36	13.67		16.92	1.75		11.86	17.22	6.43		0.09	-2.12	*	4.96
	L10%	-36.17	-3.25			-10.11			-34.44	6.34			-7.08		
CAT.N	H10%	6.49	-2.22		13.22	-8.44		0.45	13.96	10.14	*	29.84	4.34	*	17.02
	L10%	-7.34	-15.44	**		-8.90	**		-20.11	-19.70	*		-12.68		
CSCO.OQ	H10%	9.86	27.00		43.18	4.35		5.14	24.03	15.06		52.42	-2.22		9.94
	L10%	-32.26	-16.17	**		-0.79			-11.57	-37.36	***		-12.16	**	
CVX.N	H10%	6.19	-16.84		-8.22	-10.68		-6.73	11.96	2.90	*	6.94	-0.70		4.59
	L10%	-6.15	-8.62			-3.94			-23.97	-4.04			-5.29		
DD.N	H10%	2.49	-4.19		-5.28	-6.47		-7.22	2.08	-7.91		-5.32	-8.42		-4.13
	L10%	-2.92	1.09			0.76			-2.20	-2.58			-4.28		
DIS.N	H10%	7.52	5.72		8.40	4.32	*	2.54	7.83	-6.33		-6.50	-0.92		-1.03
	L10%	-3.35	-2.68			1.78			-6.32	0.18			0.10		
GE.N	H10%	8.08	-4.98		-9.96	-0.72		-6.42	22.78	16.78	***	30.82	3.82		10.84
	L10%	-37.11	4.98			5.71	*		-12.58	-14.04			-7.03		
GS.N	H10%	8.98	9.68		25.84	11.22		28.15	20.51	-7.16		-3.83	2.79		10.70
	L10%	-22.77	-16.16	*		-16.93	***		-45.40	-3.33			-7.91		
HD.N	H10%	8.83	16.87	***	10.53	2.42		3.61	19.87	12.42	**	3.45	-2.37		-9.92
	L10%	-4.21	6.34			-1.19			-7.93	8.97			7.55		
HPQ.N	H10%	9.20	16.87	**	77.73	8.00		5.80	17.54	3.83		61.17	1.80		-4.09
	L10%	-16.28	-60.86	**		2.20			-23.56	-57.33	**		5.90		

[continue table next page]

[continue of previous table]

		(1)	(2)		(3)	(4)		(5)	(6)	(7)		(8)	(9)		(10)
		$CSent_S$	$CAR0_S$	p	H-L	$CAR1_S$	p	H-L	$CSent_N$	$CAR0_N$	p	H-L	$CAR1_N$	p	H-L
IBM.N	H10%	13.19	19.13	***	61.09	10.90	**	8.69	31.87	29.72	***	70.12	18.87	***	21.15
	L10%	-11.25	-41.96	***		2.21			-17.81	-40.41	***		-2.29		
INTC.OQ	H10%	18.15	3.53		20.82	8.84		5.50	30.15	0.27		12.03	13.46	*	11.52
	L10%	-10.69	-17.30	**		3.34			-18.10	-11.76			1.94		
JNJ.N	H10%	7.26	2.86	*	16.07	-0.95		12.38	12.20	8.02	**	17.89	8.44	***	13.03
	L10%	-4.14	-13.21	**		-13.32	**		-10.04	-9.87			-4.59		
JPM.N	H10%	8.07	-4.59		-18.77	-5.17		-23.63	19.37	-6.62		-3.83	-4.95		-15.76
	L10%	-19.34	14.19			18.46	**		-39.91	-2.78			10.82		
KO.N	H10%	7.32	2.77		-3.60	-0.21		-6.90	16.61	-6.06		-0.72	-3.92		0.07
	L10%	-6.93	6.37			6.69			-12.53	-5.34			-3.99		
MCD.N	H10%	6.69	6.56		9.53	2.04		-0.11	12.62	-3.93		-3.67	-5.67	*	-6.36
	L10%	-10.30	-2.97			2.15			-19.46	-0.26			0.69		
MMM.N	H10%	5.10	-0.57		-11.33	6.01		0.49	11.16	-1.07		3.85	6.15		12.00
	L10%	-2.93	10.76	**		5.52			-7.94	-4.92			-5.85		
MRK.N	H10%	6.99	13.37	***	31.23	-2.93		9.49	10.70	13.36	**	31.77	2.63		9.53
	L10%	-5.12	-17.85	**		-12.42	*		-10.82	-18.41	***		-6.89		
MSFT.OQ	H10%	27.42	2.94		22.56	-3.08		5.45	50.42	8.45		25.48	-4.51		-0.49
	L10%	-17.56	-19.62	**		-8.53			-13.57	-17.03	**		-4.02		
NKE.N	H10%	10.28	32.03	***	128.99	-1.36		2.99	11.41	-8.83		1.20	0.50		-6.13
	L10%	-4.80	-96.96	***		-4.35	*		-8.99	-10.04			6.63		
PFE.N	H10%	7.22	14.85	*	16.71	13.16	*	6.80	12.84	8.72		17.77	1.18		4.68
	L10%	-8.05	-1.87			6.36			-14.81	-9.05	***		-3.50	**	
PG.N	H10%	4.94	8.85	*	8.80	8.75		0.51	9.57	8.64	*	12.76	2.63		4.59
	L10%	-3.24	0.05			8.24			-10.82	-4.11			-1.95		
T.N	H10%	11.18	-3.37		-0.04	-3.89		-1.81	17.87	-0.23		8.07	-3.14		1.10
	L10%	-9.85	-3.33			-2.08			-17.35	-8.30			-4.24		
TRV.N	H10%	1.48	-0.47		-10.88	-1.50		-13.92	2.30	3.03		-1.99	-6.97		-7.67
	L10%	-1.44	10.42			12.42			-1.53	5.02	**		0.70		
UNH.N	H10%	4.36	10.13		6.19	6.42		10.34	9.74	2.54		2.76	-9.47		-8.99
	L10%	-2.13	3.94			-3.92			-5.90	-0.22			-0.48		
UTX.N	H10%	3.85	-2.46		1.62	-2.50		-5.50	7.83	-1.14		0.13	-4.31		-4.98
	L10%	-3.67	-4.08			3.00			-9.22	-1.27			0.66		
V.N	H10%	6.91	17.00	*	93.77	4.81		6.99	8.05	24.37	***	11.91	4.34		-4.75
	L10%	-2.69	-76.78	*		-2.18			-4.63	12.46			9.09		
VZ.N	H10%	10.87	-6.06		6.78	-3.20		3.10	19.26	-3.86		12.04	-6.61		-1.81
	L10%	-8.83	-12.83			-6.31			-17.47	-15.90	**		-4.80		
WMT.N	H10%	11.62	-2.80		-5.83	-3.07		-6.51	20.38	-10.23		-12.50	-7.28	*	-7.85
	L10%	-15.02	3.04			3.44			-27.74	2.27			0.57		
XOM.N	H10%	7.77	-7.74		-3.76	0.89		4.52	16.46	-5.70		1.94	1.70		0.90
	L10%	-10.90	-3.98			-3.63	*		-30.43	-7.64			0.80		
Profit					**19.25**			**1.40**				**14.20**			**1.33**
Corr			**0.3997**			**0.0969**				**0.5464**			**0.1191**		

A.6 Robustness Tests

Table A.7: DOUBLE SORT: CONTROL FOR PREVIOUS DAY RETURN. This table reports double sorting results that controls for the previous day's excess returns ($CAR(t-1)$). H and L are top and bottom deciles of previous day excess returns expressed in basis points. Within the extreme deciles of each stock, H10% and L10% represent the highest and lowest deciles of sentiment. $CSent_S$ and $CSent_N$ indicate cumulative investor sentiment overnight from 16:00 on day t-1 to 9:29am on day t, in social media and news media respectively. $ACSent_S$ and $ACSent_N$ are the average cumulative investor sentiment for social media and news media respectively, by dividing cumulative sentiment of number of non-zero observations over the same time period (16:00pm-9:29am). $CAR(t)$ is the corresponding cumulative excess returns on the next day conditioning on the double sorting scheme, accumulating from 9:31am to 16:00pm. *, **, and *** denote significance level of 90%, 95%, and 99%, respectively. Confidence bands are constructed by performing bootstrap simulation repeatedly for 2,000 times, using sample size equal to the number of observations in each stock's decile bins. H-L reports the profit of a strategy that buy highest 10% sentiment days and sell lowest 10% sentiment days for each stock conditioning on good or bad previous day performance. "Profit" measures the average profits across sample stocks of longing high sentiment days and shorting low sentiment days. "Corr" represent the correlation coefficient between $ACSent_S$ and CAR(t) and $ASent_N$ and CAR(t), respectively.

RIC	CAR(t-1)			Social Media					News Media				
				$CSent_S$	$ACSent_S$	CAR(t)	p	H-L	$CSent_N$	$ACSent_N$	CAR(t)	p	H-L
AA.N	H	336.31	H10%	7.89	0.0076	-57.41			6.28	0.0066	-66.29		
			L10%	-6.48	-0.0066	57.89		-115.30	-16.86	-0.0196	-38.60		-27.68
	L	-372.17	H10%	7.15	0.0070	-5.93			7.65	0.0080	15.43		
			L10%	-10.60	-0.0103	-109.01	*	103.08	-28.83	-0.0290	-78.56		93.98
AAPL.OQ	H	202.28	H10%	47.61	0.0454	6.54			56.70	0.0541	22.87		
			L10%	**-50.31**	**-0.0480**	**-63.27**	*****	69.81	**-59.58**	**-0.0569**	**-69.18**	*****	92.06
	L	-212.00	H10%	50.03	0.0477	32.57			70.80	0.0675	39.15		
			L10%	-51.94	-0.0495	46.15		-13.58	-75.85	-0.0723	9.89		29.26
AXP.N	H	174.18	H10%	5.07	0.0051	29.31			12.89	0.0137	8.71		
			L10%	-3.90	-0.0043	-5.03		34.33	-8.70	-0.0093	-4.19		12.90
	L	-176.85	H10%	4.28	0.0044	8.04			8.71	0.0101	-60.42	**	
			L10%	-4.98	-0.0055	33.06		-25.02	-14.07	-0.0147	14.20		-74.62
BA.N	H	187.85	H10%	10.66	0.0103	18.43			34.96	0.0335	87.53	**	
			L10%	-19.35	-0.0186	-33.66	*	52.09	-28.11	-0.0270	-21.37		108.90
	L	-190.91	H10%	8.32	0.0080	-7.97			27.06	0.0260	1.57		
			L10%	-16.50	-0.0160	1.40		-9.37	-48.28	-0.0467	-11.09		12.67
BAC.N	H	266.89	H10%	8.47	0.0081	-47.45			19.57	0.0189	-6.47		
			L10%	-57.09	-0.0545	-66.37		18.92	-51.89	-0.0496	-157.28	***	150.80
	L	-300.33	H10%	6.80	0.0065	28.51			13.21	0.0129	53.75		
			L10%	-48.13	-0.0460	-127.39	***	155.90	-43.90	-0.0421	-30.70		84.45
CAT.N	H	219.96	H10%	5.27	0.0054	-20.06			8.55	0.0104	12.23		
			L10%	-7.39	-0.0077	-46.41	*	26.35	-22.84	-0.0233	14.14		-1.91
	L	-224.87	H10%	4.75	0.0049	-32.60			12.65	0.0135	6.43		
			L10%	-9.64	-0.0093	-75.64	**	43.03	-34.97	-0.0341	12.14		-5.71
CSCO.OQ	H	163.47	H10%	12.49	0.0121	23.35			23.95	0.0231	13.60		
			L10%	-25.34	-0.0242	19.97		3.38	-12.34	-0.0119	38.04		-24.44
	L	-242.89	H10%	9.93	0.0097	20.49			24.73	0.0239	24.85		
			L10%	-37.09	-0.0354	1.70		18.79	-29.94	-0.0287	25.19		-0.33
CVX.N	H	174.33	H10%	3.42	0.0035	-7.23			13.04	0.0130	-51.07	**	
			L10%	-7.95	-0.0078	-25.99		18.76	-38.59	-0.0371	8.23		-59.31
	L	-173.95	H10%	6.20	0.0062	11.99			15.34	0.0148	4.78		
			L10%	-7.26	-0.0074	-27.22		39.21	-25.99	-0.0252	11.62		-6.84
DD.N	H	185.58	H10%	2.72	0.0029	0.96			1.57	0.0480	16.20		
			L10%	-2.91	-0.0032	-15.36		16.32	-1.97	-0.0033	50.50	*	-34.30
	L	-194.48	H10%	1.99	0.0024	8.64			1.96	0.0040	48.02	*	
			L10%	-3.59	-0.0039	-51.43		60.07	-2.13	-0.0032	7.43		40.59
DIS.N	H	160.04	H10%	8.89	0.0090	24.27			5.50	0.0064	-71.93	*	
			L10%	-3.88	-0.0041	-19.82		44.10	-10.08	-0.0111	10.61		-82.54
	L	-162.57	H10%	7.50	0.0075	1.56			6.02	0.0065	32.12		
			L10%	-3.43	-0.0036	5.49		-3.93	-9.61	-0.0095	-16.97		49.09

[continue table next page]

[continue of previous table]

RIC	CAR(t-1)			Social Media					News Media				
				$CSent_S$	**$ACSent_S$**	**CAR(t)**	**p**	**H-L**	**$CSent_N$**	**$ACSent_N$**	**CAR(t)**	**p**	**H-L**
GE.N	H	169.35	H10%	5.23	0.0051	0.29			22.26	0.0214	-44.97		
			L10%	-47.52	-0.0453	-67.41	*	67.69	-14.16	-0.0136	2.97		-47.94
	L	-166.15	H10%	9.84	0.0095	-37.29			19.89	0.0194	-21.04		
			L10%	-46.20	-0.0441	28.08		-65.36	-20.42	-0.0198	21.59		-42.62
GS.N	H	209.37	H10%	7.98	0.0077	-1.88			15.18	0.0146	-10.33		
			L10%	-24.65	-0.0235	-73.41	**	71.53	-46.89	-0.0449	30.93		-41.26
	L	-215.60	H10%	5.91	0.0057	-20.10			13.62	0.0132	-10.68		
			L10%	-29.24	-0.0279	-23.63		3.54	-61.09	-0.0585	-35.37		24.70
HD.N	H	183.29	H10%	9.94	0.0098	-22.72			19.83	0.0193	20.41		
			L10%	-4.21	-0.0043	-23.87		1.15	-5.90	-0.0060	-47.96	**	68.37
	L	-176.25	H10%	8.68	0.0086	7.08			22.94	0.0223	5.37		
			L10%	-5.44	-0.0054	-90.70	***	97.78	-10.72	-0.0106	-5.11		10.48
HPQ.N	H	272.89	H10%	16.51	0.0160	-46.45			16.81	0.0163	11.93		
			L10%	-23.89	-0.0229	-38.23		-8.22	-34.09	-0.0332	53.27	*	-41.34
	L	-298.06	H10%	11.92	0.0116	35.44			22.68	0.0219	41.19		
			L10%	-21.78	-0.0209	80.60		-45.16	-41.14	-0.0398	29.41		11.78
IBM.N	H	141.25	H10%	12.67	0.0123	22.51			39.64	0.0381	33.43		
			L10%	-17.63	-0.0171	45.17	**	-22.66	-32.25	-0.0310	44.18	**	-10.75
	L	-139.59	H10%	9.20	0.0089	-2.79			33.47	0.0322	-63.34	***	
			L10%	-23.82	-0.0228	16.56		-19.35	-40.12	-0.0384	24.55		-87.89
INTC.OQ	H	190.65	H10%	21.11	0.0202	-10.39			35.35	0.0341	-26.12		
			L10%	-15.91	-0.0153	-46.91	**	36.52	-24.78	-0.0237	15.48		-41.60
	L	-160.96	H10%	20.10	0.0192	-1.61			31.25	0.0301	-24.17		
			L10%	-9.40	-0.0090	-9.92		8.31	-18.36	-0.0176	9.18		-33.35
JNJ.N	H	116.82	H10%	8.01	0.0079	9.03			24.88	0.0240	21.88		
			L10%	-4.11	-0.0043	-9.32		18.35	-8.28	-0.0084	-1.52		23.40
	L	-113.97	H10%	8.88	0.0086	6.86			9.98	0.0222	-18.60		
			L10%	-3.98	-0.0039	22.77		-15.91	-14.89	-0.0144	-2.15		-16.46
JPM.N	H	197.56	H10%	7.56	0.0074	-1.11			20.89	0.0200	-20.37		
			L10%	-21.19	-0.0203	-50.81		49.70	-49.09	-0.0470	-24.53		4.16
	L	-202.58	H10%	6.83	0.0066	8.62			15.70	0.0150	-50.59	**	
			L10%	-23.90	-0.0229	-43.04	*	51.66	-50.33	-0.0481	28.03		-78.62
KO.N	H	131.54	H10%	6.50	0.0065	4.55			15.39	0.0149	-25.48		
			L10%	-8.22	-0.0079	-18.02		22.57	-17.10	-0.0165	-29.66	*	4.18
	L	-169.89	H10%	7.22	0.0071	7.18			16.50	0.0162	8.18		
			L10%	-7.23	-0.0071	-14.82		22.00	-13.05	-0.0126	-15.29		23.47
MCD.N	H	129.70	H10%	5.20	0.0052	0.08			7.53	0.0074	26.92	**	
			L10%	-12.23	-0.0118	-26.45	*	26.53	-25.25	-0.0243	3.78		23.14
	L	-123.35	H10%	5.93	0.0058	-7.79			14.67	0.0141	-7.79		
			L10%	-11.27	-0.0109	-13.37		5.59	-25.78	-0.0247	-30.60	*	22.81
MMM.N	H	145.42	H10%	6.15	0.0079	-4.49			20.29	0.0200	2.52		
			L10%	-3.35	-0.0035	-12.82		8.33	-12.14	-0.0123	11.30		-8.78
	L	-140.07	H10%	3.68	0.0040	3.13			5.87	0.0073	46.81	*	
			L10%	-3.20	-0.0034	7.66		-4.54	-11.02	-0.0114	-63.35	***	110.16
MRK.N	H	153.34	H10%	6.14	0.0061	-0.97			12.28	0.0130	6.32		
			L10%	-4.00	-0.0040	5.56		-6.53	-5.43	-0.0082	-25.78		32.09
	L	-167.47	H10%	7.47	0.0074	15.85			8.47	0.0085	16.00		
			L10%	-6.14	-0.0060	-15.66		31.51	-20.86	-0.0201	-12.22		28.22

[continue table next page]

[continue of previous table]

RIC	CAR(t-1)			Social Media					News Media				
				$\mathbf{CSent_S}$	$\mathbf{ACSent_S}$	**CAR(t)**	**p**	**H-L**	$\mathbf{CSent_N}$	$\mathbf{ACSent_N}$	**CAR(t)**	**p**	**H-L**
MSFT.OQ			H10%	20.10	0.0192	7.90			46.68	0.0448	22.16		
	H	174.47	L10%	-19.90	-0.0192	-55.09	***	63.00	-13.81	-0.0132	-28.36		50.53
			H10%	33.25	0.0320	1.56			51.74	0.0495	-34.59		
	L	-163.90	L10%	-20.09	-0.0192	42.16		-40.60	-26.89	-0.0257	-0.78		-33.81
NKE.N			H10%	13.09	0.0154	-25.63			15.48	0.0155	-13.90		
	H	201.32	L10%	-4.79	-0.0053	50.14	**	-75.77	-7.27	-0.0073	19.99		-33.89
			H10%	13.45	0.0170	37.03			13.34	0.0133	14.74		
	L	-192.87	L10%	-2.79	-0.0028	-1.04		38.07	-6.46	-0.0073	16.26		-1.52
PFE.N			H10%	7.72	0.0075	10.77			9.87	0.0101	-18.01		
	H	155.46	L10%	-11.92	-0.0114	36.51		-25.74	-16.72	-0.0162	17.71		-35.72
			H10%	6.58	0.0064	-37.86			13.28	0.0129	3.97		
	L	-157.40	L10%	-11.16	-0.0107	65.54	**	-103.40	-19.82	-0.0196	-34.62		38.59
PG.N			H10%	4.94	0.0060	-17.39			10.58	0.0111	-4.85		
	H	123.23	L10%	-3.96	-0.0039	20.15		-37.55	-10.29	-0.0104	-7.26		2.41
			H10%	5.21	0.0063	-12.71			8.37	0.0108	-27.99		
	L	-123.70	L10%	-3.86	-0.0042	-5.30		-7.41	-16.70	-0.0165	10.61		-38.60
T.N			H10%	9.18	0.0089	1.46			16.22	0.0167	1.95		
	H	132.52	L10%	-12.42	-0.0120	-6.24		7.70	-19.47	-0.0187	21.77		-19.82
			H10%	12.28	0.0120	16.44			17.48	0.0170	4.12		
	L	-143.45	L10%	-10.77	-0.0105	-11.85		28.29	-23.62	-0.0227	1.00		3.12
TRV.N			H10%	1.31	0.0023	1.04			2.95	0.0035	-5.87		
	H	148.40	L10%	-1.61	-0.0187	-28.35		29.38	-2.19	-0.0052	-3.95		-1.92
			H10%	0.94	0.0036	-30.21			1.98	0.0024	0.49		
	L	-158.86	L10%	-1.17	-0.0021	-2.85		-27.36	-0.77	-0.0014	-6.39		6.88
UNH.N			H10%	4.88	0.0056	31.72			12.68	0.0128	-55.12	*	
	H	211.03	L10%	-4.42	-0.0044	-24.80		56.52	-7.13	-0.0172	20.35		-75.47
			H10%	2.90	0.0038	22.56			7.68	0.0091	-12.24		
	L	-200.69	L10%	-1.77	-0.0023	-4.48		27.03	-7.48	-0.0087	2.15		-14.39
UTX.N			H10%	2.60	0.0034	-5.81			10.85	0.0115	-27.82		
	H	170.51	L10%	-3.21	-0.0045	8.56		-14.37	-10.76	-0.0111	-26.08		-1.74
			H10%	3.30	0.0055	-19.95			6.26	0.0074	-40.40		
	L	-165.88	L10%	-4.20	-0.0051	-1.01		-18.94	-8.24	-0.0120	-16.93		-23.47
V.N			H10%	8.15	0.0084	19.42			11.91	0.0127	-7.21		
	H	193.99	L10%	-2.86	-0.0031	-36.08		55.51	-5.33	-0.0063	3.26		-10.47
			H10%	6.58	0.0212	13.57			9.13	0.0120	25.18		
	L	-181.41	L10%	-3.64	-0.0037	-69.38	***	82.95	-7.52	-0.0079	-42.90		68.08
VZ.N			H10%	12.18	0.0118	23.99			24.79	0.0248	18.03		
	H	151.06	L10%	-8.90	-0.0086	-21.69		45.67	-19.37	-0.0187	5.22		12.82
			H10%	10.94	0.0106	4.35			22.74	0.0218	-23.44		
	L	-153.58	L10%	-12.00	-0.0115	25.20		-20.85	-24.03	-0.0230	1.23		-24.67
WMT.N			H10%	10.13	0.0097	4.60			15.95	0.0154	19.80		
	H	140.11	L10%	-13.11	-0.0126	-22.28		26.88	-27.37	-0.0262	11.01		8.78
			H10%	15.19	0.0146	2.27			35.00	0.0336	-6.21		
	L	-147.17	L10%	-22.21	-0.0213	17.54		-15.27	-47.43	-0.0454	10.31		-16.52
XOM.N			H10%	6.86	0.0066	-0.04			18.19	0.0175	-12.88		
	H	163.95	L10%	-9.80	-0.0096	-54.21	**	54.16	-32.39	-0.0320	67.79	***	-80.67
			H10%	10.20	0.0099	29.33			17.99	0.0177	44.74	*	
	L	-145.19	L10%	-12.97	-0.0125	4.99		24.34	-37.87	-0.0364	17.04		27.69
Profit								**15.06**					**1.46**
Corr					**0.2844**					**0.0905**			