Journal of Economics and Finance (2020) 44:458–485
https://doi.org/10.1007/s12197-019-09494-4

AEF

# Stock returns and investor sentiment: textual analysis and social media

**Zachary McGurk**[1] · **Adam Nowak**[2] · **Joshua C. Hall**[2]

Published online: 3 September 2019


**Abstract**
The behavioral finance literature has found that investor sentiment has predictive ability for equity returns. This differs from standard finance theory, which provides no role for investor sentiment. We examine the relationship between investor sentiment and stock returns by employing textual analysis on social media posts. We find that our investor sentiment measure has a positive and significant effect on abnormal stock returns. These findings are consistent across a number of different models and specifications, providing further evidence against non-behavioral theories.

**Keywords** Investor sentiment · Supervised learning · Stock returns · Social media · Sufficient reduction · Predictive regression

**JEL Classification** G12 · G13 · G14

## 1 Introduction

As described in Malkiel and Fama (1970), the Efficient Market Hypothesis (EMH) predicts asset prices fully reflect all available information. Rational investors in response choose asset portfolios which diversify away idiosyncratic risk. As such

✉ Joshua C. Hall
joshua.hall@mail.wvu.edu

Zachary McGurk
mcgurkz@canisius.edu

Adam Nowak
adam.nowak@mail.wvu.edu

[1] Department of Economics & Finance, Canisus College, Buffalo, NY, USA

[2] John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA

asset prices are only a function of market fundamentals. When asset prices are mispriced through the actions of irrational investors, rational investors are able to use arbitrage to correct asset prices.

In contrast to the EMH, behavioral finance theory suggests that the feelings of irrational investors (Investor Sentiment) drive a portion of asset prices. Due to the specific characteristics of some assets (small, hard to value, limited information, etc.), arbitrage by rational investors becomes costly and asset prices are perpetually mispriced.[1] Recent empirical studies have found Investor Sentiment to be related to stock returns.[2]

While the empirical finance literature has found Investor Sentiment to be a valid predictor of the cross section and time series of stock returns, studies differ how the Investor Sentiment measure is estimated. As noted by Baker and Wurgler (2006, 2007) Investor Sentiment is difficult to directly measure. As a result, the literature has relied on proxies developed from market/investor surveys, data mining methods, and textual analysis from annual reports, commercial media, and social media.

Due to data limitations, the market/investor survey and data mining methods literature focus on the impact of investor sentiment on returns over monthly or larger time horizons. While most of these studies show a relationship between asset returns and investor sentiment, these studies may not capture the full impact of investor sentiment. If asset markets are partially efficient (i.e. investor sentiment does not determine a portion of stocks), and information is randomly dispersed, then markets should be the least efficient in the very short run.

Another critique of this literature is that these investor sentiment measures show overall market sentiment rather than asset specific sentiment. Baker and Wurgler (2006) discusses that due to imperfect information about smaller firms, any new information causes investors to engage in irrational speculative trading. Market sentiment may not necessarily capture this speculative feeling in smaller firms. While much of the textual analysis literature has been able to account for the preceding critiques, the estimation methods used may be not be able to fully capture investor sentiment. A portion of the previous literature has relied on dictionary based methods in determining Investor Sentiment (Loughran and McDonald 2011; Chen et al. 2014; Jiang et al. 2019). To estimate sentiment, these studies pre-define a dictionary of positive and negative finance words and determine overall investor sentiment as the net positive word counts. The limitation of this approach is that there may be important missing terms which show sentiment. This method also gives each word equal weight in determining sentiment and does not account for sentiment shown in multi-word phrases.[3] Other studies have utilized machine learning methods to estimate investor sentiment (Bartov et al. 2018; Ranco et al. 2015; Yang et al. 2015; Sun et al. 2016; Renault 2017; Behrendt and Schmidt 2018). These studies provide an improvement on dictionary-based methods as the created investor sentiment indexes

[1] See, for example, Baker and Wurgler (2006) and Baker and Wurgler (2007).

[2] See Bukovina (2016) and Zhou (2018) for a review of the recent literature.

[3] Loughran and McDonald (2011) include a method for weighting individual words, however, this is based on word frequency rather than perceived sentiment information.

allow for different weighting of textual terms. These papers focus on the extreme short run (5 to 30-minute intervals) impact of investor sentiment on returns and given data limitations are unable to create equity specific investor sentiment.[4]

Given the limitations of the previous literature, we propose a new method for estimating Twitter based stock specific investor sentiment index utilizing as developed in Taddy (2013a). This method differs in that estimates of sentiment do not rely on a predefined dictionary, and individual words are not assumed to be related the same sentiment information. Further, given the data rich environment of Twitter, we are able to create equity specific investor sentiment indexes. In this method, a training set of posts by individual users on Twitter (tweets) are determined to either convey positive, neutral, or negative sentiment. These are then used to predict the sentiment information from all remaining tweets. For comparison, we also develop a dictionary based investor sentiment utilizing a similar method as Loughran and McDonald (2011).

We further utilize our investor sentiment index to test the empirical validity of EMH and Behavioral Finance theories. We specifically determine the relationship between our investor sentiment measures (negative, neutral, and positive sentiment) and cross-section abnormal stock returns. For robustness, we test if this relationship is similar across firm size. Finally, we determine if investor sentiment is useful in forecasting abnormal returns at the market level and by firm size.

The social media platform Twitter is used by over 320 million users who express opinions and thoughts on a number of different subject matters including equity prices.[5] Further, Twitter is unique in that an individual can reference specific stocks by affixing a '$' before the stock symbol in a tweet. This allows all Twitter users to search for tweets discussing a particular stock. This allows researchers to collect tweets supplied by individuals specific to a stock. Anecdotal evidence has shown individual Twitter posts (tweets) to influence specific stock returns. On January 10, 2011, *Business Insider* reported hip hop artist, 50-Cent (Curtis Jackson), tweeted

> HNHI is the stock symbol for TVG there launching 15 different products. they are no joke get in now.

The article goes on to state (Weisenthal 2011, no page number):

> In the three months to the end of September, the company was operating at a loss with cash of just $198,000 and a deficit of $3.3m. Then, on November 23, it said it would offer 180m shares to the public at a price of just 17 cents... trading under the stock name HNHI was worth just 4 cents each. Spurred by the tweet, the stock took off. It hit nearly 50 cents on Monday, before closing at 39 cents.

[4]Bloomberg and Thompson Reuters have created commercial equity specific textual analysis based investor sentiment measures. These measures are proprietary and as such estimation methods are unknown. These measure are used by Sun et al. (2016) and Behrendt and Schmidt (2018).

[5]Source: Twitter 2018 Annual Report.

By the end of the month, the stock was up to $1.68. This price increase was relatively short lived. In early May, 50-cent terminated his relationship with this company, and the stock dropped in value to $0.1.

Overall, we find a relationship between abnormal stock returns and our estimated investor sentiment indexes. We find an increase in positive sentiment is related to an increase in abnormal returns while also finding that estimated negative estimated sentiment had a limited relationship with abnormal returns. These results are consistent across firm size. Using out-of-sample forecasting tests, we find investor sentiment is able to produce marginally more accurate forecasts compared to a constant only model. Gains in forecast accuracy, however, is limited to around one percent. Our results indicate that individuals on Twitter are relaying stale information as opposed to providing novel insights.

The remainder of the paper proceeds as follows. Section 2 details the relevant literature on investor sentiment and textual analysis, Section 3 describes the methodology and data, Section 4 details the cross sectional analysis, Section 5 provides a discussion on forecasting method utilized and forecasting results, and Section 6 concludes.

## 2 Literature review

The empirical finance literature has overwhelmingly shown a relationship between investor sentiment and cross-sectional stock returns. Since (Baker and Wurgler 2006, 2007), the primary focus of the literature has been to determine an appropriate proxy for investor sentiment. Specifically, the literature has focused on three potential sources of investor sentiment proxies: investor and consumer surveys, methods similar to Baker and Wurgler (2006) such as Huang et al. (2015), and textual analysis from traditional and social media.

The investor sentiment proxies derived from surveys and using methods similar to Baker and Wurgler (2006) and Huang et al. (2015) have generally found that investor sentiment is related to cross section and future returns (Lee et al. 2002; Brown and Cliff 2004; Lemmon and Portniaguina 2006; Smales 2017; Shen et al. 2017). Smales (2017) finds that the Chicago Board Options Exchange Volatility Index (VIX) performs empirically well compared to survey measures. Baker and Wurgler (2006) and Huang et al. (2015) find the same for stock returns. Chung et al. (2012) using a (Baker and Wurgler 2006) style investor sentiment index, finds that sentiment is unable to forecast returns during recessions. In a related work, Shen et al. (2017) find that the difference in returns from portfolios sorted by macroeconomic risk are related to the Baker and Wurgler (2006) investor sentiment index. Aboody et al. (2018) suggest overnight returns may be an appropriate proxy for investor sentiment and find that high overnight returns predict returns.

More recently, studies have focused on the impact of investor sentiment in international financial markets. Utilizing consumer confidence as a measure of investor sentiment, Schmeling (2009) finds that sentiment can forecast market returns for 18 industrialized countries. Baker et al. (2012) show that global investor sentiment is able to predict market returns for six developed countries. Frijns et al. (2017) show

that US investor sentiment (as measured by the American Association of Individual Investors Investor sentiment survey) is related to market returns for several developed countries. Seok et al. (2018) find a high frequency firm specific investor sentiment index utilizing a Baker and Wurgler (2006) method for the Korean stock market can predict future returns in the Korean stock market. Xu and Zhou (2018) utilizing an investor sentiment based on Huang et al. (2015) find market investor sentiment can predict future returns in the Chinese A-share market.

The use of surveys and methods similar to Baker and Wurgler (2006) and Huang et al. (2015) as a measure of investor sentiment has remained controversial. Da et al. (2015) note that the investor sentiment surveys are generally unreliable given there is little incentive for survey takers to respond or respond truthfully.

The investor sentiment proxies created using surveys and methods similar to Baker and Wurgler (2006) and Huang et al. (2015) are at best available weekly but often less frequently. Further, these measures proxy for overall market sentiment. If a portion of the market is efficient, then it is likely the shorter the time horizon the larger the impact of investor sentiment on the cross section of stock returns. With longer time horizons, rational investors may be able use arbitrage to counter the mispricing caused by irrational investors. In contrast, with the amount and frequency of information available on an ongoing basis, internet and social media derived indexes can produce higher frequency and stock specific measures of investor sentiment.

While overall market-level investor sentiment is likely driving the systemic mispricing of assets, equity-specific investor sentiment is likely to play a role in idiosyncratic mispricing. Further, as discussed in Baker and Wurgler (2007), investor sentiment is likely to have a different impact on pricing based on asset characteristics, leading to "safe" assets being undervalued during periods of positive sentiment and overvalued during periods of negative sentiment.

We argue that while market sentiment plays a role in explaining the overall mispricing of assets, stock-specific sentiment is likely to be informative. Avery and Chevalier (1999) provides three potential sources of sentiment bias in sports betting markets: so-called expert opinions, a hot-handed bias, and a bias toward prestigious teams. Avery and Chevalier (1999) further note these sources have their equivalent in asset markets. These sources may lead retail investors to speculate about specific firms regardless of their overall feeling toward asset markets.

With the frequency and availability of firm specific information provided in traditional and social media, textual analysis provides an avenue for further study of investor sentiment. Two procedures have been used in the textual analysis literature. Both of these treat the text as a collection of exchangeable *tokens*, a token being either a word or a phrase. The first approach is the so-called *bag-of-words* approach that requires the researcher to specify a dictionary of positive and negative tokens (Loughran and McDonald 2011, 2014; Chen et al. 2014; Heston and Sinha 2017; Renault 2017; Jiang et al. 2019). The second approach, which we refer to as the *tokenization* approach, does not require the researcher to explicitly specify any prior beliefs regarding the positivity or negativity of individual tokens but rather uses manually labeled text to identify relevant tokens (Taddy 2013a, b; Mitra and Gilbert 2014; Ranco et al. 2015; Oliveira et al. 2016; Renault 2017).

The literature has primarily relied on two distinct word lists to develop dictionaries using the *bag-of-words* approach: Harvard IV-4 word lists,[6] and the Loughran and McDonald (2011) finance specific word list. The Harvard IV-4 word list was create to determine the tone of texts for the sociology and psychology literature. This dictionary is further divided into several different categories with the negative and positive word lists getting the largest use in the finance literature.

Loughran and McDonald (2011) argue that the Harvard IV-4 word list is not applicable to finance texts and can lead to the misidentification of sentiment. Loughran and McDonald (2011) argue that the Harvard IV-4 word list does not include a number of key finance specific tokens. Further, certain tokens are misclassified as negative when used in a finance context. In fact, Loughran and McDonald (2011) find that around 74 percent of the negative tokens found in the Harvard IV-4 word list are not deemed negative in a finance context. To address these critiques, Loughran and McDonald (2011) create a finance specific word list to accurately identify the tone of 10-K filings. Loughran and McDonald (2011) find sentiment of 10-K filings using their finance specific dictionaries are more correlated with equity returns compared to sentiment of 10-K filings using the Harvard IV-4 word list. Chen et al. (2014) utilize the Loughran and McDonald (2011) dictionary to create an equity specific investor sentiment index from posts and comments from Seeking Alpha (a crowd-sourced financial market media source). Chen et al. (2014) find their investor sentiment index is able to predict stock returns up to three months ahead. Jiang et al. (2019) utilize the Loughran and McDonald (2011) dictionary to create a manager sentiment index from annual and quarterly filing, and conference calls. Overall (Jiang et al. 2019) find that manager sentiment is able to predict stock returns using out-sample forecast evaluation.

The *bag-of-words* method, while being thoroughly used in the literature, has several limitations. First, the *bag-of-words* method relies on the researcher to correctly identify an appropriate dictionary prior to determining the tone of a text. Loughran and McDonald (2016) note that the use of the Loughran and McDonald (2011) finance specific dictionary for use in anything but determining the tone of 10k filings can be problematic and modification must be done to identify sentiment from other forms of media. Creating a finance specific dictionary or modifying the Loughran and McDonald (2011) dictionary using only one word tokens (unigrams) that are applicable to specific media is a doable task. Attempting to do this for dictionaries utilizing multi-word tokens (n-grams) may be improbable, as tens of thousands tokens would need to be identified. This dictionary would not just include tokens and their antonyms (e.g., Sell and Don't Sell), but other tokens that may not found in unigrams (e.g., Death Cross). While not directly using the *bag-of-words* approach, Yang et al. (2015) provides a limited solution to this by modifying estimated sentiment from a tweet when a token with a negative connotation precedes another token in their dictionary.

Even given these critiques, *bag-of-words* methods may be harder to implement when estimating sentiment from social media. Loughran and McDonald (2016) state

[6] http://www.wjh.harvard.edu/∼inquirer/homecat.htm

that given the specific vocabulary used in social media (e.g., slang, and sarcasm), correctly identifying tone is much more difficult compared to SEC filings. Second, a majority of the *bag-of-words* literature has relied on counts of positive and negative tokens in text or other methods of equal weighting of tokens in a dictionary to determine sentiment.[7] This implies that each token provides the same amount of sentiment information. It is likely that some tokens present greater negative or positive sentiment than others. Loughran and McDonald (2011) suggests weighting tokens by frequency to account for the differences in document size. Further, when longer texts are written by fewer authors, *bag-of-words* estimated sentiment without token weighting may lead to an incorrect identification of sentiment due to the number of words. Specifically, authors may differ in how they express their emotions (overtly or minimally) through text and as such estimated sentiment may be biased by the number of tokens.[8]

An advantage of using posts from Twitter to estimate sentiment is that our sample contains posts from many users. This reduces the likelihood that our estimation suffers from small sample bias caused by users expressing their emotions differently. Further, as the number of possible characters are limited to 140 during our sample period, this weighting method is likely not necessary to infer sentiment from Twitter.[9]

The *tokenization* approach provides a solution to many of the above critiques. Specifically, the *tokenization* approach does not require an author to predefine a dictionary. Further, methods like the Multinomial Inverse Regression (MNIR), developed in Taddy (2013b), allow for each token to be weighted by the sentiment information that it provides. As described in Loughran and McDonald (2016), the inverse regression method can be viewed a method for "identifying and weighting sentiment words."

In the context of creating an investor sentiment index, the standard procedure for implementing the *tokenization* approach is to create a training set where a sample of text is manually labeled as either positive, neutral, or negative.[10] From this scored set of texts, the researcher can: 1) identify which tokens are associated with positive, neutral or negative texts, 2) estimate weights that indicate the relative positivity or negativity associated with each token, and 3) use the predicted values to create sentiment scores for each document.

A number of studies utilize a similar *tokenization* method, Naïve Bayesian classifiers, to identify investor sentiment from social media. Bartov et al. (2018) use both a *tokenization* and *bag-of-words* approach to develop equity specific investor sentiment index from Twitter. Bartov et al. (2018) finds their Twitter sentiment can predict a stock's return prior to earnings announcements. With the *tokenization* approach, individual tweets are weighted by their estimated probability of being negative, positive,

[7] See Loughran and McDonald (2016) for an extended discussion of alternative weighting schemes.

[8] We thank the anonymous reviewer for providing an additional critique of the *bag-of-words* approach without token weighting.

[9] In 2017, Twitter increased character limits to 280. As the twitter data used in study is from prior to this change, all tweets are limited to 140 characters.

[10] Alternative approaches can be taken that would allow for continuous classifications.

or neutral and by the number of followers of the Twitter user. Ranco et al. (2015) creates an investor sentiment index for 30 stocks in the DJIA using Twitter and tests if large tweet increases relate to future abnormal returns. Ranco et al. (2015) find estimated negative and positive sentiment Granger cause abnormal returns. Yang et al. (2015) estimate a twitter sentiment index using only tweets from user considered "influencers." The general idea is that influencers are the ones who are showing market sentiment while other users are simply trying to find information to trade on. The investor sentiment index is created by using the "comprehensive dictionary" (SentiWordNet). While the SentiWordNet dictionary is derived using a machine learning method, it was not created for a specifically for use in finance and its use may be suspect when applied to a finance-based application. Yang et al. (2015) find past sentiment is related to current market returns.

Several recent studies have focused on determining the role of investor sentiment in determining and predicting intraday returns. Oliveira et al. (2016) and Renault (2017) utilize both a *tokenization* and *bag-of-words* approach to create an investor sentiment index using posts from StockTwit.[11] This social media platform is unique in that users can pre-label their tweets as either "Bullish" (demonstrating negative sentiment) or "Bearish" (demonstrating positive sentiment). This allows the researcher to forgo manually labeling a training set. This study shows the StockTwit derived investor sentiment indexes can forecast intraday S&P 500 EFT returns. Sun et al. (2016) find that changes in the Thomson Reuters sentiment index predict market returns with a larger impact on the last two hours of trading. In a similar study, Behrendt and Schmidt (2018) utilizes the Bloomberg News and Social Sentiment index to determine if changes in this index can predict intraday stock returns. This study finds that while the investor sentiment measure is statistically significant it is not economically significant. As discussed in Renault (2017), utilizing these measures of investor sentiment are hard to judge given that they are proprietary and such are not "transparent, or replicable."

Given their short time horizons (30 minutes for Sun et al. (2016) and Renault (2017) and five minutes for Behrendt and Schmidt (2018)), these studies are unable to create stock specific investor sentiment indexes. It is further likely the estimated investor sentiment measures are biased toward sentiment information on larger and well known firms. If the theory presented in Baker and Wurgler (2006, 2007) is correct, the impact of investor sentiment may be under estimated.

The *tokenization* approach requires an indicator variable for each token. As such, the number of variables can grow large for even a modest number of observations which necessitates the need for dimension reduction. Given this, we estimate our Twitter based investor sentiment utilizing MNIR for two specific reasons. First, MNIR is a dimension reduction method that links token frequencies to class labels. Exploiting this linkage contrasts with principal components analysis, which would not use information in labels but only information contained in the covariance matrix of indicator variables. Second, MNIR uses sparsity-inducing priors in which many coefficients are set equal to 0. By doing so, the MNIR selects a set of empirically

[11] StockTwit is social media platform where by users primarily discuss topics relating to financial markets.

relevant tokens. This proves incredibly beneficial for the *bag-of-words* approach as results from MNIR can be used to build dictionaries or validate existing dictionaries. The sparse solution of MNIR contrasts with principal components analysis where factors are estimated as linear of combinations of all tokens. As noted in the literature, interpreting such loadings can be incredibly difficult.

To gauge investor sentiment, we aggregate information from 2.5 million tweets that explicitly reference a stock symbol. Each tweet contains character data that must be processed into a numeric variable that measures sentiment.

## 3 Methodology and data

The goal of any textual analysis is to accurately score a given body of text. The following section details the procedures we use to score the individual tweets and create a stock specific investor sentiment index. We specifically create two sets of investor sentiment indexes employing both the *tokenization* approach of Taddy (2013b) and *bag-of-words* approach.

### 3.1 Tokenization approach

Each tweet contains 140 characters or less. Each tweet can be indexed by date, $t$, and company, $j$. On each date, there are $i = 1, ..., N_{jt}$ tweets that discuss company $j$. To refer to a specific stock, Twitter users include $ and the stocks ticker symbol. For instance, when discussing Exxon (Ticker symbol: XOM) and Haliburton (Ticker symbol: HAL) an individual will use *$HAL* in the tweet. A figurative example of standard tweet in our sample is provide below.

$$sold\ \$XOM\ shares\ and\ bought\ \$HAL\ shares \tag{1}$$

In the following discussion, we abstract from $ijt$ subscripts and instead use $n = 1, ..., N$ when referring to any of the $N$ tweets. Each tweet, $x_n$, contains tokens that describe sentiment regarding company $j$. That is, we can describe $x_n$ as a set of $W_n$ tokens. When the tokens are partitioned by unigrams, $x_n$ is written as:

$$x_n = \{w_n^1, w_n^2, ..., w_n^{W_n}\} \tag{2}$$

Following the previous example, we have:

$$x_n = \{sold, \$XOM, shares, and, bought, \$HAL, shares\} \tag{3}$$

Similar to the literature, we employ the following procedures to process each tweet and retain only possible token information.[12]

- Convert all text to lowercase characters.

[12] We utilize the MNIR method to first obtain a set of tweets relevant information. Further information on this method can be provided upon request.

- Remove all stop words. Stop words are common words that do not directly indicate sentiment including prepositions, articles, pronouns, etc.[13]
- Remaining words are stemmed to retain a root form of the word. For example, *buying, bought, buy* are converted to *buy*.
- Convert the stock symbol of the reference company to *COMPANY* and all other stock symbols in the tweet to *OTHERCOMPANY*. By doing so, we can properly associate the sentiment with the stock.
- Drop all numbers from the tweet.[14]

Following the above procedure, if the reference company is Haliburton, the example tweet, would now read:

$$x_n = \{sold, OTHERCOMPANY, shares, bought, COMPANY, shares\} \quad (4)$$

From the example tweet, it is clear that the user has a positive outlook on Haliburton and a negative outlook on Exxon. The example tweet highlights the importance of word ordering in determining sentiment from tweets. For instance, if we were to switch *sold* and *bought* it would appear as if the individual is bearish on Exxon and bullish on Haliburton. Using the *bag-of-words* approach, it is likely that this tweet would be scored as neutral.

Word order is further important when a negator precedes a word that expresses either positive or negative sentiment. For example, if a tweet read "Do not Sell Haliburton," partitioning the tweet by unigram would not account for the positive sentiment expressed from "not Sell."

A simple way to control for ordering is to use two word phrases, bigrams, as the tokens.[15] By partitioning the example tweet into bigrams, $x_n$ becomes:

$$x_n = \{sold.OTHERCOMPANY, OTHERCOMPANY.shares, shares.bought, bought.COMPANY, COMPANY.shares\} \quad (5)$$

### 3.2 Multinomial distribution

We use a collection of $k = 1, ..., K$ unique tokens collected from the $N$ tweets. The $W_n$ tokens in $x_n$ are represented as draws from a multinomial distribution with unknown token probabilities. In other words, we draw $W_n$ tokens from $K$ possible tokens, where each token is drawn with probability $q_{nk}$. When using the multinomial distribution, we assume the tokens in $x_n$ are exchangeable.

For each token, define the indicator variable $z_n^k = 1$ if token $k$ is in the set $x_n$ and $z_n^k = 0$ otherwise. Associated with these indicator variables is the vector $z_n$, the $K \times 1$

[13] Because the tweets frequently indicate the direction of the stock (up or down), we modify a list of stop words from the *SnowballC* package in R to retain finance-specific words.

[14] In unreported results, we found our quantitative results were not substantially impacted by this procedure. However, the list of significant tokens is impacted by this procedure, as expected.

[15] We use the convention where different words in a bigram are separated by "."

vector of these indicator variables. We assume each tweet can be classified into one of 3 distinct, unordered categories: $positive$, $negative$ or $neutral$. Without loss of generality, define the class of each tweet as $y_n \in \{1, 2, 3\}$. In what follows, it will be useful to keep track of $y_n$ by defining the $3 \times 1$ vector $v_n$ where the $y_n$ element is equal to 1, and all other elements are equal to 0.

This classification scheme is not exhaustive. There exist an innumerable number of ways to classify each tweet based on various criteria. It should be emphasized that, although the categories appear to have an inherent ordering, the procedure used in this study does not require such an ordering. Further, we only classify tweets into the $positive$ and $negative$ categories. Because of this, the $neutral$ category might also be thought of as a $neither$ category.

Conditional on $y_n$ and $W_n$, we assume that $x_n$ is drawn from a multinomial distribution. Using notation similar to Taddy (2013b), the vector of token counts, $z_n$, is distributed as a multinomial random variable $z_n \sim MN(q_n, W_n)$ where:

$$\begin{aligned} q_{nk} &= \frac{e^{\eta_{nk}}}{\sum_l e^{\eta_{nl}}} \\ \eta_{nk} &= \alpha_k + v_n \psi_k + \epsilon_{nk} \end{aligned} \tag{6}$$

In Eq. 6 conditional on $y_n$, $q_{nk}$ indicates the probability that token $k$ will be in $x_n$. $\alpha_k$ is a token specific parameter. When token $k$ appears more frequently across all classes, $\alpha_k$ will be large. $\psi_k$ is a $3 \times 1$ vector of parameters specific to token $k$ that indicates the class-specific relative frequency of token $k$. The product $v_n \psi_k$ is the $y_n$ element of $\psi_k$; the $y^{th}$ element of $\psi_k$ will be positive whenever token $k$ appears more frequently in tweets of class $y$ compared to other classes. The term $\epsilon_{ij}$ is an i.i.d. error term.

Given the parameters $\psi_k$, we can summarize all relevant information in $x_n$ using the linear combination $s_n^T = W_n^{-1} \Psi' z_n$ where $\Psi' = [\psi_1, ..., \psi_K]$. Taddy (2013b) shows that this linear combination is a sufficient reduction where by the distribution of $y_n$ given the $3 \times 1$ vector $s_n^T$ is independent of the distribution of $y_n$ given the entire $K \times 1$ vector $z_n$. For the purposes at hand, $s_n^T$ can be interpreted as a measure of sentiment. As such, six investor sentiment indexes are created; positive, neutral, and negative for both unigrams and bigrams.

In a regression context, $s_n^T$ can be used as an explanatory variable in a parsimonious regression in lieu of using the potentially thousands of regressors in $z_n^T$. At its core, the above procedure is a method to reduce the dimension of the token vector $z_n$.

As mentioned above, the total number of unique tokens can be quite large. Any model with a large number of parameters runs the risk of over-fitting. In our estimation, we experimented with various token sets and settled on a modest cutoff using only the 3,000 most frequent tokens when estimating the model. In unreported results, the coefficient estimates were robust to token sets as large as 5,000. To prevent this, Taddy (2013b) uses a Laplace prior for the $\psi_k$ and selects the vector $\widehat{\Psi}$ as the vector which maximizes the posterior likelihood given the priors. Due to Laplace priors being used, the resulting $\widehat{\Psi}$ is sparse with many elements equal to 0. In this sense, the procedure is both variable selection and coefficient estimation. From a

predictive standpoint, only tokens with non-zero coefficients in $\widehat{\Psi}$ are relevant when predicting the class of the tweet.

### 3.3 Supervised learning

To estimate Eq. 6, it is necessary to use a training set of tweets labeled as either negative, neutral or positive. Taddy (2013b) uses restaurant ratings labeled by multiple users. Taddy (2013a) in measuring political sentiment towards presidential use both emoticons and the Amazon Mechanical Turk service to read and score tweets.[16]

Labeling tweets that reference stocks is not as straightforward as labeling tweets that reference political candidates. Most tweets manifest a clear sentiment. Examples include *$XOM looks like a buy* and *bearish on $XOM*. However, due to esoteric finance vocabulary, it is possible that a randomly selected individual will miss subtleties associated with payoffs that can lead to an incorrect labeling. Such instances frequently occur when discussing options contracts. For example, *bought May puts* and *bought May calls* should be classified as negative and positive, respectively, although this distinction is not apparent to an individual without an understanding of finance terminology. Taddy (2013a) takes several steps to pre-screen the workers to ensure the fidelity of their labeling scheme. However, due to cost considerations, we concluded that using the Amazon Mechanical Turk service was not viable option.

Instead, we manually label a random sample of 3000 tweets as a training set. Similar to the arbitrary notions expressed in Baker and Wurgler (2006) regarding the measurement of sentiment, it is possible that our categories and labeling procedure are also arbitrary and subjective. To mitigate such problems, we undertook several steps to produce results that were both objective and thorough. First, a minimum observation period of 30 seconds was required before scoring was allowed to ensure a sufficient amount of time was spent analyzing the tweet. Second, a repeated sampling procedure was performed to ensure consistent labeling by author. Third, some tweets were scored by each author to mitigate any individual bias. Any tweets that were had multiple labels by the same individual or across individuals were discarded. This resulted in a collection of 2919 labeled tweets or an error rate of approximately 2.7%. An examination of these discarded tweets may indicate human error and not any discrepancy in the perceived sentiment of the tweet. The training set is then used to estimate $s_{ijt}^{T}$, the estimated sentiment of the tweet.

Table 1 shows the largest coefficients estimated for the unigram and bigram methods for both positive and negative sentiment. Panel A shows the coefficients of the tokens most related to positive sentiment. For both unigrams and bigrams the tokens, "buy" and "long" are estimated to be highly related with positive sentiment. Surprisingly, the tokens "green" and "biotechmoney" are found to have a strong relationship with positive sentiment. These are non-finance specific token which may be picking up positive sentiment towards the biotech and environmental industry. Panel B shows

[16] See https://www.mturk.com/ for information on the Amazon Mechanical Turk service.

**Table 1** Estimated token coefficients

| Unigram | | Bigram | |
|---|---|---|---|
| Panel A: Positive | | | |
| highs | 1.1 | buy COMPANY | 1.3 |
| buy | 1 | long COMPANY | 1.1 |
| long | 1 | biotechmoney COMPANY | 1 |
| added | 0.9 | look OTHERCOMPANY | 1 |
| positive | 0.8 | special dividend | 1 |
| green | 0.7 | call options | 0.9 |
| strong | 0.7 | eye COMPANY | 0.9 |
| dividend | 0.6 | look COMPANY | 0.9 |
| breaking | 0.5 | looks good | 0.9 |
| Panel B: Negative | | | |
| lower | 0.7 | morning COMPANY | 1.4 |
| weak | 0.7 | taking COMPANY | 1.4 |
| profits | 0.8 | downgrades COMPANY | 1.5 |
| seeing | 0.7 | sell half | 1.5 |
| bearish | 1.3 | PRICE puts | 1.8 |
| short | 1.4 | short term | 1.8 |
| downgrades | 1.5 | sell COMPANY | 1.9 |
| sell | 1.5 | short COMPANY | 1.9 |
| shorted | 1.7 | drops COMPANY | 2 |

Note: Table 1 shows the largest estimated token coefficients using the *tokenization* approach. Panel A shows the largest estimated coefficients of tokens both unigram and bigram predicting positive sentiment. Panel B shows the largest estimated coefficients of tokens both unigram and bigram predicting negative sentiment. A larger estimated coefficient relates to a higher estimated probability the tweet is classified as positive or negative. Estimates of token coefficients for predicting neutral sentiment are not shown for brevity

the estimated coefficients of the tokens for negative sentiment. Unlike the positive sentiment, these tokens are primarily finance words or phrases which correspond to negative sentiment. Tokens like "short," "sell," "downgrades," and "weak" all seem to indicate negative feelings towards assets.

Table 2 provides a list of the ten most negative and most positive tweets based on $s_{ijt}^{T}$. The original tweet is shown where the reference stock symbol has been removed and replaced with *COMPANY*. For many of these tweets, the user is primarily stating their position (long or short) with respect to a specific asset.

### 3.4 Bag-of-words approach

While the focus of this study is to develop an investor sentiment index using the *tokenization* approach of Taddy (2013b), we utilize the *bag-of-words* method to create

**Table 2** Examples of labeled positive and negative tweets: supervised method

| |
|---|
| Panel A: Positive tweets |
| 1: rt @reddogt3live: nice to see the treasure for trash trade inching higher. still long some COMPANY as it still has upside room. entries and exits matter |
| 2: rt @redacre: COMPANY upgraded by mizuho securities from neutral to buy @chasingthealpha see? #pumpers - COMPANY was a screaming buy at $25 ; below |
| 3: bought COMPANY $92.5 jan calls...looks like its about to break |
| 4: pr: special dividend declaration alert for COMPANY (expect long). |
| 5: bought COMPANY at $3.55 and then again at $2.74. sitting over $9 now. #letitride |
| 6: rt @sciencetrader: things i like with good long term charts that are high but have room to run: $xlb, $itb, $xhb, COMPANY, $vz, $oih, $lmt, $wfm,( $amzn needs time) |
| 7: bought some COMPANY and $vvus today |
| 8: paper bought 5000 COMPANY jan 2014 40 calls for $.88 |
| 9: bought COMPANY jan 2014 95 calls for 1.33 #optionsaction |
| 10: @jimcramer hi j, bought COMPANY jan 600 calls in oct. when stock was at 610. down now, time decay a worry, do i ride to the end? |
| Panel B: Negative tweets |
| 1: i am still short COMPANY and $pkx. getting tough to stay short as the major averages and sector indexes are oversold near term. |
| 2: sold some COMPANY march 85 calls for $0.73 |
| 3: sold COMPANY stock at 628.50 for +6.00 – hedge now removed. 20 sma test ; rejection was the trigger. |
| 4: good time to reset the portfolio...sold COMPANY and $csco holdings...time to find the next investments #retirement #dividends |
| 5: sold all COMPANY for $1.5k profit. what an amazing move today. |
| 6: sold those COMPANY jan calls, +16.66%, ffiv, -5.2%. |
| 7: sold COMPANY at $4.87 from $4.80 entry. small gain |
| 8: sold my COMPANY mar $5 calls at $1.2 (from $1.4) - no need to be a hero holding thru briefing docs tomorrow |
| 9: thinking of shorting COMPANY ... just keeps going down |
| 10: sold more COMPANY +60% |

Note: Table 2 provides a snapshot of the ten most negative and most positive tweets estimated using the bigram *tokenization* approach. Note prior to scoring, tweets have been processed using the steps described in Section 3.1

a comparable index. Following the critique of Loughran and McDonald (2016), we create a Twitter finance dictionary of positive and negative unigrams we believe Twitter users would use when referring to stocks. This dictionary is built using unigram counts from the labeled data from the procedure above. We further divide our dictionary into a Positive Word List (PWL) and Negative Word List (NWL). The PWL is a list of words or phrases that might indicate the Twitter user is positive about the stock. The NWL is likewise a list of words or phrases that might indicate negative sentiment.

**Table 3** Finance dictionary

| |
|---|
| Panel A: Positive dictionary |
| buy, buying, bought, long, bull, bullish , good, acceptable, excellent, exceptional, favorable, great, positive, awesome |
| Panel B: Negative dictionary |
| added, sell, selling, sold, short, bear, bearish, bad, atrocious, awful, crummy, dreadful, lousy, poor, rough, sad, unacceptable, blah, bummer, downer, gare, gross, imperfect, inferior, junky, abominable, amiss, crappy, cruddy, dissatisfactory, erroneous, fallacious, faulty, godawful, inadequate, substandard, unsatisfactory, shitty |

Note: Table 3 shows positive and negative finance dictionary used in estimating the *bag-of-words* investor sentiment measure

While the literature has primarily used either the Harvard IV-4 word list or a unmodified (Loughran and McDonald 2011) finance, we argue this would be inappropriate in determining sentiment from Twitter given the relatively small character limit and informal language used on social media. Loughran and McDonald (2016) states "The use of slang, emoji, and sarcasm, and the constantly changing vocabulary on social media makes the accurate classification of tone difficult." With this in mind, a dictionary containing a large number of tokens would likely perform as well as a smaller dictionary. Further utilizing a larger dictionary may lead to overfitting. As such we favor a dictionary with fewer tokens. Our Twitter finance dictionary is given in Table 3, with the list of positive and negative unigrams. These tokens can indicate either a company fundamentals, trading positions, or technical indicators.

$PWL$ and $NWL$ are shown as mathematical sets below:

$$PWL = (PWL_1, PWL_2, ...) = (bought, bullish, breakout, ...) \tag{7}$$

$$NWL = (NWL_1, NWL_2, ...) = (old, bearish, dismal, ...) \tag{8}$$

Using $PWL$ and $NWL$, three measures of sentiment are estimated. First, positive sentiment is the total count of positive words in each tweet. Positive sentiment for a given tweet is written as:

$$Ps_n^B = \#(x_n \cap PWL) \tag{9}$$

Here, $\#(X)$ is the number of words in the set $X$. Second, negative sentiment is the total count of negative words in each tweet. Third, overall sentiment is estimated positive sentiment subtracted by negative sentiment. Negative sentiment for a given tweet is written as:

$$Ns_n^B = \#(x_n \cap NWL) \tag{10}$$

We determine overall sentiment of each tweet by counting the total number of words in $PWL$ minus the total number of words in $NWL$. The sentiment for a given tweet is written as:

$$s_n^B = \#(x_n \cap PWL) - \#(x_n \cap NWL) \tag{11}$$

If there are more positive words than negative words in $x_n$, then $s_n^B > 0$ and vice-versa.

Differences and similarities between the *bag-of-words* approach and the *tokenization* approach can be found by comparing $s_n^B$ and $s_n^T$. Each measure of sentiment is a linear combination of the vector $z_n$. However, for the *bag-of-words* approach, positive tokens are given a weight $+1$ and negative tokens are given a $-1$. In the *tokenization* approach, tokens are assigned weights based on the strength of their association with negative and positive tweets. Of course, it is possible to estimate weights for tokens in the dictionary using Eq. 6 or by specifying weights ex-ante. However, we find that the words in the dictionary are almost always selected when using a much larger set of tokens.

For comparison between *bag-of-words* and the *tokenization* approach, we also estimate the 5 most negative and positive tweets for $s_n^B$, found in Table 4. By comparing

**Table 4** Examples of labeled positive and negative tweets: *bag-of-words*

| |
|---|
| Panel A: Positive tweets |
| 1: @JEFFREYJKEITH no mans land when in doubt buy time. I'm going long cover me . Jan 600/700 bull call spread at 52.5 COMPANY |
| 2: Nervous abt buying COMPANY at 545ish then buy deep money long dated calls. Too expensive? Then buy call spread (lower cost+limit prof) |
| 3: Very tempted to buy some COMPANY... havent bought a non-financial long time. Other than $YNGFF.yuck |
| 4: Not 2 long ago, everyone was saying COMPANY was a buy at $550 and change; goin to $600, then $1,000.00 millions of people bought . They lost. |
| 5: Everyone has been saying buy COMPANY and shit, well based on my charting skills, and bearish look on $GOOG now is time to long COMPANY 430.58 |
| Panel B: Negative tweets |
| 1: sold $tza jul $11 ns contracts for 1.04, basis .84. added to COMPANY jul short $6 puts by selling .65. added to $wlt, sold $cprx |
| 2: also added to COMPANY am & sold 1/2 at 6.38 for (avg breakeven) & pondered last 15 mins before selling other 1/2 now at 6.52 for nice gain. |
| 3: @tlmontana @mwonder74 i had sold my COMPANY shares & replaces them w/ short puts over $34. i think dips buyable. sell otm puts. |
| 4: path as i did with COMPANY .. best way sell to open, sold jan $5 calls for $0.25 ... will keep premium or short it at $5 on opex |
| 5: positions. added to $npcuf. holding $exel, exel nov calls, small $ziop & $ttnp. short $osir. sold COMPANY on bounce; will re-enter after nce. |

Note: Table 4 shows provides a snapshot of the most negative and most positive tweets based on the *bag-of-words* method. Note prior to scoring, tweets have been processed using the steps described in Section 3.1

Tables 2 and 4, the *tokenization* approach anecdotally seems able to more accurately determine the correct tone of tweets. For example, the *bag-of-words* approach seems to mislabels the 2nd most positive tweet as positive. This tweet is more likely displaying negative sentiment. The tweet reads:

$$\textit{Nervous abt buying COMPANY...} \tag{12}$$

which implies the investor is unsure the price of that stock will increase.

### 3.5 Timing and stock returns

Currently, the New York Stock Exchange (NYSE) is open for trading business days between 9:30 AM and 4:00 PM EST; in contrast social media users can post content 24 hours a day/7 days a week. The focus of our analysis is determining if information from social media can explain variation in abnormal stock returns. If the standard 24 hour convention is used to estimate daily sentiment, a portion of tweets would be posted after the market has closed. Thus possibly causing any relationship found between investor sentiment and abnormal returns to spurious. Because of this, is necessary to be precise when indicating the start and end of a period. We use two breakpoints in order to split the 24 hour day into three segments. The two breakpoints indicate the opening and closing of the NYSE. The day is then split up into three periods defined by:

- Pre-Market: 00:00:00-09:29:59
- Market: 09:30:00-15:59:59
- Post-Market: 14:00:00-23:59:59

We regress abnormal returns on the sentiment calculated within each time segment. That is, we regress abnormal return on pre-market sentiment, market sentiment and post-market sentiment. We are interested in whether or not sentiment is correlated with current abnormal stock returns. Using our three investor sentiment indexes created using the *tokenization* approach we estimate the following equation:

$$AR_{jt} = \alpha + \gamma_1 Negative_{jt} + \gamma_2 Neutral_{jt} + \gamma_3 Positive_{jt} + u_{jt} \tag{13}$$

Following Sprenger et al. (2014a, b), $AR_{jt}$ is the abnormal return of the security calculated as the raw return of the stock $j$ at time $t$ minus the return on the S&P 500 at time $t$. $Negative_{ijt}$, $Neutral_{ijt}$, and $Positive_{ijt}$ represent all of the elements in $s^T_{ijt}$. $Negative_{jt}$ is estimated as the average negative sentiment for each stock $j$ at day $t$ calculated by summing the individual $Negative_{ijt}$ across all $i = 1, ..., N_{jt}$ and dividing by $N_{jt}$. $Neutral_{ijt}$ is estimated as the average neutral sentiment for each stock $j$ at day $t$ calculated by summing the individual $Neutral_{ijt}$ across all $i = 1, ..., N_{jt}$ and dividing by $N_{jt}$. $Positive_{jt}$ is estimated as the average positive sentiment for each stock $j$ at day $t$ calculated by summing the individual $Positive_{jt}$ across all $i = 1, ..., N_{jt}$ and dividing by $N_{jt}$.

Using the investor sentiment index created with the *bag-of-words* approach, we estimate the following equation:

$$AR_{jt} = \alpha + \beta_1 Dictionary_{jt} + u_{jt} \tag{14}$$

As in Eq. 13, $AR_{jt}$ is the abnormal return of the security calculated as the raw return of the stock $j$ at time $t$ minus the return on the S&P 500 at time $t$. $Dictionary_{jt}$ is the estimated average net positive sentiment for each stock $j$ at day $t$ calculated by summing the individual $Dictionary_{ijt}$ across all $i = 1, ..., N_{jt}$ and dividing by $N_{jt}$ using the *bag-of-words* approach.

### 3.6 Data and Descriptive Statistics

The Twitter data was collected using the *TwitteR* package in R beginning August 20, 2012, and ending June 12, 2013. Data was collected daily by searching Twitter for tweets discussing stocks in the Russell 5000. The search program allowed us to search for the 5,000 most recent tweets from the past seven days for any number of stocks. At the start of the data collection, Twitter allowed for an unlimited number of searches. However, by mid-2013, Twitter limit the number of downloads, and data collection was effectively terminated. Over this period, we collected 3,941,149 unique tweets over 296 days discussing 4,972 unique stocks.

Table 5 shows descriptive statistics for the investor sentiment measures, returns, and firm characteristics. The number of total observations is 116,649. This counts for the total number of sentiment measures for each firm (which were tweeted about) on average for each day. This relates to on average 23 days per firm and 792 tweets for each firm. This also corresponds to around 13,314 tweets about firms per day.

**Table 5** Descriptive statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| $\text{Negative}_{Uni}$ | 116,649 | −0.482 | 0.218 | −2.623 | −0.301 |
| $\text{Neutral}_{Uni}$ | 116,649 | −0.103 | 0.150 | −2.037 | −0.0001 |
| $\text{Positive}_{Uni}$ | 116,649 | −1.596 | 0.388 | −2.365 | 1.819 |
| $\text{Negative}_{Bi}$ | 116,649 | −0.449 | 0.170 | −2.205 | −0.340 |
| $\text{Neutral}_{Bi}$ | 116,649 | −0.095 | 0.131 | −1.778 | −0.0001 |
| $\text{Positive}_{Bi}$ | 116,649 | −1.576 | 0.270 | −1.696 | 1.036 |
| Dictionary | 116,649 | 1.873 | 15.921 | −2,989 | 1,156 |
| Price | 116,649 | 38.826 | 58.048 | −123.235 | 2,766.000 |
| Volume | 116,649 | 4,031,242 | 10,589,024 | 0 | 463,491,000 |
| Return | 116,649 | 0.403 | 3.756 | −49.624 | 49.383 |
| Book Value | 116,649 | 7,584.627 | 22,426.350 | −8,341.000 | 236,956.000 |
| Market Value | 116,649 | 15,463.110 | 39,799.860 | 0.018 | 499,821.000 |

Note: Table 5 shows the descriptive statistics for a number of financial variables and investor sentiment measures. The first column shows the variables. Here $_{Uni}$ represents unigram estimates and $_{Bi}$ represents bigram estimates. The second column shows the number of observations for each variable. The third column shows the estimated mean. The fourth column shows the estimated standard deviation of each variable. The fifth column and sixth column shows the minimum and maximum value for each variable respectively. The top half shows the descriptive statistics for the estimated sentiment measures. The bottom half shows the descriptive statistics for the individual firm characteristics. Only descriptive statistics for the All timing are shown

The *bag-of-words* investor sentiment measure, *Dictionary*, has a mean of 1.9 which corresponds to an average positive sentiment during this period. While descriptive statistics are presented for the all *tokenization* approach investor sentiment measures, as the literature notes these values are relatively difficult to interpret. We do find that unigram means are higher in absolute terms than bigram means. Further, the standard deviations for all bigrams are lower than for unigrams. This may be caused by bigrams providing a more precise and accurate estimate of investor sentiment.

Firm characteristics include Trade Volume, Price, Return, Book Value and Market Value. The mean daily Return is 0.4 percent with a minimum and maximum -49.6 and 49.4 percent.[17] Market value is calculated by multiplying total shares by price. Market value is shown in millions. The mean market value is about 15 billion with a standard deviation of 39 billion. As our sample covers approximately 99 percent of US issued stocks, the range of market value is relatively large. The smallest firm has a market value of 18 thousand and the largest 499 billion. In comparison, Book Value is relatively smaller with mean value of 7 billion and a range from -8 billion to 237 billion.

## 4 Cross-sectional results

Using ordinary least squares (OLS), we estimate Eq. 13 for both the unigram and bigram investor sentiment measures. Tables 6 and 7 show the cross sectional market level results using the estimated sentiment from the *tokenization* approach. Table 8 shows the OLS results for Eq. 14 using sentiment estimated from the *bag-of-words* approach.

We overall find strong empirical evidence that our estimated investor sentiment indexes are related to abnormal returns. These results are consistent with much of the previous literature and provides additional evidence against the efficient market hypothesis in favor of behavioral finance theory.

Table 6 shows the cross sectional results utilizing the unigram sentiment measures. We find that higher Negative, Neutral, and Positive sentiment are in general related to higher abnormal returns. The coefficient for the positive unigram measure is significant for all timings, while the Negative and Neutral sentiment measures are significant for all but the Pre-Market timing. The results here are similar to the results found for the bigram sentiment measures, found in Table 7. Both the coefficients for the positive and neutral bigram measures are significant for at least three of the timings. In contrast to the unigram results, the coefficient of the negative bigram measure is only significant for Pre-Market and Post-Market timing although only at the 10% significance level.

In a surprising result, we find the coefficients for negative unigram sentiment statistically significant and positive. This result implies that with an increase in negative sentiment (more negative sentiment) is related to higher abnormal returns. These results are in contrast to the negative bigram results which find that the coefficient

[17] To limit the effect of outliers, we only include daily Returns in between -50 and 50 percent.

**Table 6** Cross sectional results - *Tokenization Approach*: Unigram sentiment

| | *Dependent variable: Abnormal Returns* | | | |
|---|---|---|---|---|
| | All | Pre-Market | Market | Post-Market |
| Negative | 0.691*** | 0.191 | 0.624*** | 0.822*** |
| | (0.071) | (0.127) | (0.108) | (0.090) |
| Neutral | 0.684*** | −0.109 | 0.483*** | 0.682*** |
| | (0.079) | (0.134) | (0.116) | (0.105) |
| Positive | 0.621*** | 0.254*** | 0.546*** | 0.641*** |
| | (0.038) | (0.067) | (0.057) | (0.049) |
| Observations | 116,649 | 34,342 | 62,740 | 76,778 |
| $R^2$ | 0.003 | 0.001 | 0.002 | 0.003 |
| F Statistic | 111.363*** | 6.526*** | 36.217*** | 74.722*** |

Note: Table 6 shows the cross sectional results from unigram estimated sentiment. Specifically, OLS is used to estimate Eq. 13. All, Pre-Market, Market, and Post-Market represent results using the timing conventions described in Section 3.5. Negative represents estimated negative sentiment using the *tokenization* approach. Neutral represents estimated neutral sentiment using the *tokenization* approach. Positive represents estimated positive sentiment using the *tokenization* approach. Estimated coefficient for each are shown. Standard errors are presented in parenthesis. Constant term is omitted for brevity. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level

**Table 7** Cross sectional results - *Tokenization Approach*: Bigram sentiment

| | *Dependent variable: Abnormal Returns* | | | |
|---|---|---|---|---|
| | All | Pre-Market | Market | Post-Market |
| Negative | 0.120 | 0.326* | 0.115 | 0.168* |
| | (0.080) | (0.168) | (0.116) | (0.099) |
| Neutral | 0.797*** | −0.215 | 0.582*** | 0.794*** |
| | (0.088) | (0.152) | (0.127) | (0.116) |
| Positive | 0.308*** | 0.326*** | 0.360*** | 0.177*** |
| | (0.048) | (0.104) | (0.067) | (0.064) |
| Observations | 116,649 | 34,342 | 62,740 | 76,778 |
| $R^2$ | 0.001 | 0.0004 | 0.001 | 0.001 |
| F Statistic | 40.246*** | 4.036*** | 15.983*** | 21.486*** |

Note: Table 7 shows the cross sectional results from bigram estimated sentiment. Specifically, OLS is used to estimate Eq. 13. All, Pre-Market, Market, and Post-Market represent results using the timing conventions described in Section 3.5. Negative represents estimated negative sentiment using the *tokenization* approach. Neutral represents estimated neutral sentiment using the *tokenization* approach. Positive represents estimated positive sentiment using the *tokenization* approach. Estimated coefficient for each are shown. Standard errors are presented in parenthesis. Constant term is omitted for brevity. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level

**Table 8** Cross sectional results - *Bag-of-Words Approach*

| | *Dependent variable: Abnormal Returns* | | | |
|---|---|---|---|---|
| | All | Pre-Market | Market | Post-Market |
| Dictionary | 0.003*** | 0.001 | 0.005*** | 0.006*** |
| | (0.001) | (0.002) | (0.002) | (0.001) |
| Observations | 116,649 | 34,342 | 62,740 | 76,778 |
| $R^2$ | 0.0002 | 0.00001 | 0.0001 | 0.0003 |
| F Statistic | 20.224*** | 0.348 | 8.979*** | 19.407*** |

Note: Table 8 shows the cross sectional results from *bag-of-words* approach. Specifically, OLS is used to estimate Eq. 14. All, Pre-Market, Market, and Post-Market represent results using the timing conventions described in Section 3.5. Dictionary represents the estimated net positive sentiment measure using the *bag-of-words* approach. Estimated coefficients for Dictionary are shown. Standard errors are presented in parenthesis. Constant term is omitted for brevity. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level

estimates are positive and significant at the 10 % level but only for the pre-market and post-market timing. Given the critiques of unigram estimates as compared to bigram or other n-gram estimates, as described in Section 2, the negative unigram measure may not be correctly measuring negative sentiment. Anecdotally, as shown in Table 1, the token "profits" is found to have a large impact on unigram negative sentiment. The inclusion of the token "profits" by itself may not be expressing negative sentiment unless an additional token like "lower" was previously included. In fact, for the bigram estimates, "profits COMPANY" is found to have no impact on estimated negative sentiment.[18]

Table 8 shows the estimated results utilizing the *bag-of-words* approach. We find the estimated coefficient of dictionary sentiment to be positive and significant for all timings except Pre-Market. Specifically, we find the more net positive estimated sentiment is related to higher abnormal returns.

While all investor sentiment estimates perform relatively well for at the All, Market, and Post-Market timing, these measures generally perform poorly for Pre-Market timing. Specifically, we find only the coefficients of the Positive bigram and unigram measures are significant at the 1 % level for the Pre-Market timing. As such higher positive sentiment prior to the opening of the market are related to higher abnormal returns during that day.

These results are general consistent with the previous literature (Bartov et al. 2018; Ranco et al. 2015; Oliveira et al. 2016; Renault 2017), which finds that social media based investor sentiment is related to a cross section of stock returns. To the best of our knowledge, our study is the only once which is able to differentiate the impact of negative and positive sentiment on returns. Our results, from the bigram investor sentiment estimates, highlight positive sentiment maybe more important than negative sentiment in determining the cross section of abnormal returns.

[18] Bigram and Unigram coefficient estimates are available upon request.

### 4.1 Robustness by size

Baker and Wurgler (2006, 2007) detail that due to limited information, smaller and less-known stocks are particularly influenced by investor sentiment. The idea is that any information released about these stocks will cause investors to speculate and drive abnormal returns. This would imply the relationship between abnormal returns and investor sentiment should be stronger for smaller firms. To test this, we divide the sample by decile market capitalization and estimate Eq. 13 for both the unigram and bigram investor sentiment measures and Eq. 14 for the Dictionary sentiment measure.[19] We do not report the results here due to space limitations.[20]

Overall, while we find evidence that investor sentiment is related to abnormal returns for larger firms, we do not find similar evidence for lower deciles. Very few investor sentiment measures are significantly related to abnormal returns for the first four deciles. These results does not necessarily provide empirical evidence against the theory that smaller firms are disproportionally affected by investor sentiment (Baker and Wurgler 2006, 2007), as sample sizes are relatively small. Fewer tweets are available for smaller firms, so individual tweets have larger weight. A few extremely positive or negative tweets can determine this relationship. Further, many smaller firms have zero tweets for a number of days. This results in small sample sizes for the first four deciles, ranging from around 200 to 3300.

Another interesting result is that the magnitude of all estimated coefficients decrease as size increases. This implies that abnormal returns for medium sized firms are more affected by changes in investor sentiment compared to larger firms. It may be that medium sized firms are large enough that Twitter users are actively tweeting about those stocks but small enough that the impact of their tweets on speculation is heightened. These results are similar to Seok et al. (2018), who find a similar decrease in coefficients with higher deciles. Smales (2017) does not find this pattern when comparing small and large cap portfolios. Ni et al. (2015) and Yang and Zhou (2016) find the impact of investor sentiment is consistent across firm size for the Chinese stock market. Clearly, this remains an area where more empirical research is needed.

Similar to the market level results, the negative unigram measure performs better across all deciles compared to the negative bigram measure. Again, it is likely that the negative unigram measure is not fully measuring negative sentiment. The coefficient for the negative bigram measure is only found to be significant for second and tenth decile but with different signs. This provides further empirical evidence of negative sentiment having limited impact on abnormal returns.

For both the bigram and unigram, positive sentiment performs relatively well, with significant coefficients across most deciles. All significant coefficients estimates are positive. The dictionary results are relatively inconsistent across deciles. With positive and significant coefficients for only the second, fifth, sixth, ninth and tenth

[19] As a note, due to the change in the daily value of market capitalization, some stocks would be part of a different decile at different dates.

[20] Interested readers can find them in the working paper version of this article, available at https://researchrepository.wvu.edu/econ_working-papers/.

decile. These results are consistent with the market level results. As such we find higher positive sentiment relating to higher abnormal returns.

## 5 Forecasting

While we find results similar to the previous literature that investor sentiment is an important determinant of a cross section of abnormal returns. Another aim of this paper has been to determine if our investor sentiment measure can produce more accurate forecasts of returns. To do this, we employ both in-sample and out-of-sample analysis to determine the marginal gain in forecast accuracy from the inclusion of investor sentiment measures.

The forecasting literature has found investor sentiment useful in forecasting returns. Yang et al. (2015) estimate a model with just returns and one lag of sentiment at the daily level and finds the twitter-based investor sentiment measure is able to forecast several market indexes.

In intraday forecasting analysis, Sun et al. (2016) and Renault (2017) estimate a model with lags of intraday change in sentiment. Sun et al. (2016) determine if the lag of the change investor sentiment can predict the returns from half hour returns from the S&P 500 ETF. Sun et al. (2016) find the change in investor sentiment is able to predict returns up to six hours ahead. Renault (2017) utilizes a model including the change from the previous day sentiment to the first half hour of investor sentiment and half hour lags for the 11, 12, and 13 half hours in the trading day. Renault (2017) finds that only a few measures are able to predict returns of the S&P 500 ETF. For robustness a forecasting model including lag returns is estimated and does not improve forecast accuracy over other models.

Jiang et al. (2019) determine if current manager sentiment is able to predict cumulative returns from 1, 3, 6, 9, 12, 24, and 26 months ahead. All sentiment measures perform relatively well with the biggest gain being 9 months ahead. Xu and Zhou (2018) estimate a model with the lag of investor sentiment and current values of Fama French 3 factors. Lagged sentiment is able to more accurately forecast the returns for a number of size weighted portfolios. Heston and Sinha (2017) use lags of news, positive and negative sentiment to forecast returns. Positive and negative sentiment is able to forecast returns under a number of weekly forecast horizons.

Following the literature (Yang et al. 2015; Sun et al. 2016; Heston and Sinha 2017; Renault 2017; Jiang et al. 2019), we estimate the following forecasting model utilizing the unigram and bigram investor sentiment indexes:[21]

$$AR_{j,t+1} = \alpha + \gamma_1 Negative_{j,t} + \gamma_2 Neutral_{j,t} + \gamma_3 Positive_{j,t} + u_{j,t+1} \quad (15)$$

where Negative, Neutral, and Positive are the current value of the estimated investor sentiment indexes. AR represents that the future value of the abnormal return.

[21] Alternative model specifications including additional variables are also estimated. Results are relatively similar to the results from Eq. 15. These are available upon request. Due to data limitations, we only use a one day forecast horizon.

For the *bag-of-words* estimated index we estimate the following forecasting model:[22]

$$AR_{j,t+1} = \alpha + \gamma_1 Dictionary_{j,t} + u_{j,t+1} \quad (16)$$

where Dictionary represents the current value of the investor sentiment estimated through the *bag-of-words* approach and AR represents abnormal returns.

The out-of-sample analysis is a forecasting technique where the sample is split into two portions by time. The first portion is used to estimate the parameters. The estimated parameters are then used to forecast the second portion. The out-of-sample analysis allows the researcher to determine how well the model forecasts in the past. To estimate forecast accuracy the ratio of the mean squared forecast errors (MSFE) between a constant only model and estimated Eqs. 15 or 16 is taken. When the MSFE ratio is less than one, we can say that Eqs. 15 or 16 produces a more accurate forecast.

In contrast, in-sample analysis takes advantage of the full data set to estimate the relationship over the whole period. A F-test or t-test is used to determine if the additional investor sentiment measures can provide added predictability.

There has been considerable debate in the literature regarding which method (in-sample or out-of-sample) is superior in determining the relatively forecast ability of empirical models. In general, this literature has focused on the cases when the results from in-sample and out-of-sample differ (Inoue and Kilian 2005; Clark 2004, 2005). These theoretical works show conflicting evidence, with Clark (2004, 2005) finding evidence in favor of out-of-sample analysis and Inoue and Kilian (2005) finding evidence in favor of in-sample analysis. We only focus on results which match in-sample and out-of-sample.

### 5.1 Forecasting results

Table 9 shows the market level in-sample and out-of-sample forecasting results for the one day ahead forecast. For the in-sample analysis, t-test and F-test are used to determine added forecast accuracy over a constant only model. For out-of-sample analysis, the MSFE ratio is shown. For out-of-sample analysis, a sample split date of February 3, 2013 is used. This relates to a split sample ratio of around 0.9. Given the limited difference in timing only the all sample timing results are shown.

Overall, we find that twitter-based investor sentiment can improve forecast accuracy. We find the models containing the unigram and bigram investor sentiment measures are jointly significant using F-tests on the in-sample analysis. The Dictionary measure is not significant. In out-of-sample analysis, all three models are able to produce more accurate forecasts compared to a constant only model. Out of all of these models, the bigram investor sentiment produces the largest gain in out-of-sample forecast accuracy of 1.1 percent. Unigram investor sentiment improves forecast accuracy by 1.0 percent.

[22] We would like to thank the anonymous reviewer for noting non-linear forecasting models, as employed in Bekiros et al. (2016), would likely produce larger gains in forecast accuracy compared to a linear model for both approaches. Including a non-linear forecasting model in our paper does not allow for a direct comparison between our *tokenization* approach and the previous literature so we do not pursue it here. However, this is an excellent idea for future papers.

**Table 9** Forecasting results - market level

| | *Dependent variable: Abnormal Returns* | | |
|---|---|---|---|
| | Unigram | Bigram | Dictionary |
| Negative | 0.337*** | 0.177** | |
| | (0.075) | (0.080) | |
| Neutral | −0.428*** | −0.442*** | |
| | (0.083) | (0.092) | |
| Positive | 0.126*** | 0.061 | |
| | (0.044) | (0.046) | |
| Dictionary | | | 0.001 |
| | | | (0.001) |
| Observations | 113,108 | 113,108 | 113,108 |
| $R^2$ | 0.0003 | 0.0002 | 0.00001 |
| Out-of-sample MSFE Ratio | 0.990 | 0.989 | 0.999 |
| F Statistic | 12.663*** | 8.453*** | 0.624 |

Note: Table 9 shows the market level forecasting results for both sentiment estimated using the *tokenization* and *bag-of-words* approaches. The second column shows the forecasting results from estimated unigram sentiment. The third column shows the forecasting results from estimated bigram sentiment. For these first two, OLS is used to estimate Eq. 15. The fourth column show the forecasting results from estimated sentiment using the *bag-of-words* approach. For this measure, OLS is used to estimate Eq. 16. Only the All timing conventions described in Section 3.5 is shown. Negative represents estimated lag of negative sentiment using the *tokenization* approach. Neutral represents estimated lag of neutral sentiment using the *tokenization* approach. Positive represents estimated lag of positive sentiment using the *tokenization* approach. Estimated coefficient for each are shown. Standard errors are presented in parenthesis. In-sample analysis results are given by the F-statistic. Out-of-sample analysis results are given by the Out-of-sample MSFE Ratio. Constant term is omitted for brevity. *** denotes 1% significance level, ** denotes 5% significance level and * denotes the 10% significance level

For the unigram model, the in-sample t-test shows that all three sentiment measures are individually able to forecast abnormal returns. This contrasts with the bigram model, where only Negative and Neutral sentiment are found to be able to forecast abnormal returns.

The in-sample and out-of-sample results only match for both the unigram and bigram model. This gives some evidence that investor sentiment can predict future abnormal returns. Given the large volatility of stock returns/abnormal market returns, these results do not imply that the unigram and bigram model can produce relatively accurate daily forecasts. It does mean that it can produce more accurate forecasts compared to a model with a constant. Overall, we find evidence that not only is investor sentiment an important determinant of abnormal returns, it is also able to predict abnormal returns.

### 5.2 Forecasting by market capitalization

Continuing with the idea that smaller firms may be asymmetrically effected by investor sentiment provided by Baker and Wurgler (2006), we determine how well

our investor sentiment measures can forecast abnormal returns by market capitalization deciles. We again use in-sample and out-of-sample analysis and focus on the matching results. We summarize our results here and the full results are available in the working paper version of our paper.[23] Overall, unlike the market level results we are unable to find relative gains in forecast accuracy from models including investor sentiment compared to a constant only model.

Using in-sample analysis, we find that the coefficients on the unigram investor sentiment measures are jointly significant for the second, fourth, and seven deciles. Using out-of-sample analysis, we find improved forecast accuracy only for the fourth and seventh decile. The bigram model results differ in that we do not find any of the models producing more accurate forecasts, either in-sample or out-of-sample. We find the coefficients on all investor sentiment measures being jointly significant for the second, third, and fourth deciles, while in our out-of-sample analysis, we only find an improvement in forecast accuracy of 0.01 percent for the fifth decile. The *bag-of-words* results show the coefficient of the dictionary measure is significant for third through fifth and tenth deciles, using in-sample testing. The out-of-sample MSFE ratio are below one for only the third, fifth, and tenth decile, showing an increase in forecast accuracy. The out-of-sample gains in forecast accuracy range from 0.1-0.4%, with the largest gains for the 3rd decile.

The most interesting part of these results are that each investor sentiment measure differs in which decile it produces significant results and by gains in forecast accuracy. The gains may seem rather small, but due to this being daily forecasts, these are substantial over monthly or quarterly horizons. While not being able to produce relatively better results compared to for cross sectional or market level forecasting results, we find evidence of the *bag-of-words* approach producing the most accurate forecast for decile abnormal returns compared to other measures.

## 6 Conclusion

Standard finance theory predicts abnormal returns should not be a function of any variable. This result breaks down when investors do not have perfect information and act irrationally. There has been little empirical evidence that these assumptions hold. The behavioral finance has explained mis-pricing of assets through investor sentiment. We overall find evidence of investor sentiment playing a role in determining the cross section of abnormal stock returns.

We estimate a daily, firm-specific, investor sentiment measure. We improve upon the *bag-of-words* method, which gives equal weight to each word, by using supervised sentiment measures. We find that overall investor sentiment is driving overall cross section of abnormal returns with both methods. This relationship is strongest for the smallest, and therefore, least known firms. We take this as being additional evidence for small firms being particularly vulnerable to investor sentiment.

[23] Available at https://researchrepository.wvu.edu/econ_working-papers/.

We also find limited evidence that investor sentiment can produce more accurate forecasts compared to a constant only model. Increases in forecast accuracy for the overall market may be relatively small but over longer time horizons these may be substantial. Further gains are found depending on the size. With the increased use of social media, more information on the individual feeling of investors will become available. This should allow future researchers to understand the long run implications of individual firm investor sentiment. With complete Twitter data, future research can test the role of each of the sources of investor sentiment (Avery and Chevalier 1999) on abnormal returns. Specifically, the number of followers or retweets may be useful in determining the effect of expert opinion caused investor sentiment on abnormal returns. Future research should also look to employ non-linear forecasting models such as employed in Bekiros et al. (2016).

## References

Aboody D, Even-Tov O, Lehavy R, Trueman B (2018) Overnight returns and firm-specific investor sentiment. J Financ Quant Anal 53(2):485–505

Avery C, Chevalier J (1999) Identifying investor sentiment from price paths: The case of football betting. J Bus 72(4):493–520

Baker M, Wurgler J (2006) Investor sentiment and the cross-section of stock returns. J Financ 61(4):1645–1680

Baker M, Wurgler J (2007) Investor sentiment in the stock market. J Econ Perspect 21(2):129–152

Baker M, Wurgler J, Yuan Y (2012) Global, local, and contagious investor sentiment. J Financ Econ 104(2):272–287

Bartov E, Faurel L, Mohanram PS (2018) Can twitter help predict firm-level earnings and stock returns? Accounti Rev 93(3):25–57

Behrendt S, Schmidt A (2018) The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. J Bank Financ 96:355–367

Bekiros S, Gupta R, Kyei C (2016) A non-linear approach for predicting stock returns and volatility with the use of investor sentiment indices. Appl Econ 48(31):2895–2898

Brown GW, Cliff MT (2004) Investor sentiment and the near-term stock market. J Empir Financ 11(1):1–27

Bukovina J (2016) Social media big data and capital markets – an overview. J Behav Exper Financ 11:18–26

Chen H, De P, Hu YJ, Hwang B.-H. (2014) Wisdom of crowds: The value of stock opinions transmitted through social media. Rev Financ Stud 27(5):1367–1403

Chung S-L, Hung C-H, Yeh C-Y (2012) When does investor sentiment predict stock returns? J Empir Financ 19(2):217–240

Clark TE (2004) Can out-of-sample forecast comparisons help prevent overfitting? J Forecast 23(2):115–139

Clark TE, McCracken MW (2005) The power of tests of predictive ability in the presence of structural breaks. J Econ 124(1):1–31

Da Z, Engelberg J, Gao P (2015) The sum of all FEARS investor sentiment and asset prices. Rev Financ Stud 28(1):1–32

Frijns B, Verschoor WF, Zwinkels RC (2017) Excess stock return comovements and the role of investor sentiment. J Int Financ MarketInstit Money 49:74–87

Heston SL, Sinha NR (2017) News vs sentiment: Predicting stock returns from news stories. Financ Anal J 73(3):67–83

Huang D, Jiang F, Tu J, Zhou G (2015) Investor sentiment aligned: a powerful predictor of stock returns. Rev Financ Stud 28(3):791–837

Inoue A, Kilian L (2005) In-sample or out-of-sample tests of predictability: Which one should we use? Econ Rev 23(4):371–402

Jiang F, Lee J, Martin X, Zhou G (2019) Manager sentiment and stock returns. J Financ Econ 132(1):126–149
Lee WY, Jiang CX, Indro DC (2002) Stock market volatility, excess returns, and the role of investor sentiment. J Bank Financ 26(12):2277–2299
Lemmon M, Portniaguina E (2006) Consumer confidence and asset prices: Some empirical evidence. Rev Financ Stud 19(4):1499–1529
Loughran T, McDonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. J Financ 66(1):35–65
Loughran T, McDonald B (2014) Measuring readability in financial disclosures. J Financ 69(4): 1643–1671
Loughran T, McDonald B (2016) Textual analysis in accounting and finance: a survey. J Account Res 54(4):1187–1230
Malkiel BG, Fama EF (1970) Efficient capital markets: a review of theory and empirical work. J Financ 25(2):383–417
Mitra T, Gilbert E (2014) The language that gets people to give: Phrases that predict success on Kickstarter. In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. ACM, pp 49–61
Ni Z-X, Wang D-Z, Xue W-J (2015) Investor sentiment and its nonlinear effect on stock returns – new evidence from the Chinese stock market based on panel quantile regression model. Econ Model 50:266–274
Oliveira N, Cortez P, Areal N (2016) Stock market sentiment lexicon acquisition using microblogging data and statistical measures. Decis Support Syst 85:62–73
Ranco G, Aleksovski D, Caldarelli G, Grčar M, Mozetič I (2015) The effects of Twitter sentiment on stock price returns. PloS One 10(9):e0138441
Renault T (2017) Intraday online investor sentiment and return patterns in the US stock market. J Bank Financ 84:25–40
Schmeling M (2009) Investor sentiment and stock returns: Some international evidence. J Empir Financ 16(3):394–408
Seok SI, Cho H, Ryu D (2018) Firm-specific investor sentiment and daily stock returns. The North American Journal of Economics and Finance
Shen J, Yu J, Zhao S (2017) Investor sentiment and economic forces. J Monet Econ 86:1–21
Smales LA (2017) The importance of fear: Investor sentiment and stock market returns. Appl Econ 49(34):3395–3421
Sprenger TO, Sandner PG, Tumasjan A, Welpe IM (2014a) News or noise? Using Twitter to identify and understand company-specific news flow. J Bus Financ Account 41(7-8):791–830
Sprenger TO, Tumasjan A, Sandner PG, Welpe IM (2014b) Tweets and trades: The information content of stock microblogs. Eur Financ Manag 20(5):926–957
Sun L, Najand M, Shen J (2016) Stock return predictability and investor sentiment: a high-frequency perspective. J Bank Financ 73:147–164
Taddy M (2013a) Measuring political sentiment on Twitter: Factor optimal design for multinomial inverse regression. Technometrics 55(4):415–425
Taddy M (2013b) Multinomial inverse regression for text analysis. J Am Stat Assoc 108(503):755–770
Weisenthal J (2011) 50 Cent's tweets make a staggering $50 million in one day. Business Insider, 10 January
Xu H-C, Zhou W-X (2018) A weekly sentiment index and the cross-section of stock returns. Financ Res Lett 27:135–139
Yang SY, Mo SYK, Liu A (2015) Twitter financial community sentiment and its predictive relationship to stock market movement. Quant Financ 15(10):1637–1656
Yang C, Zhou L (2016) Individual stock crowded trades, individual stock investor sentiment and excess returns. North Amer J Econ Financ 38:39–53
Zhou G (2018) Measuring investor sentiment. Ann Rev Financ Econ 10:239–259