HOW MANY WORDS IS A PICTURE WORTH?

USING EMOJIS FROM SOCIAL MEDIA TO PREDICT FUTURE STOCK RETURNS*

Corbin A. Fox
James Madison University
fox2ca@jmu.edu

Eric K. Kelley
University of Tennessee
ekk@utk.edu

Roman Paolucci
Columbia University
roman.paolucci@columbia.edu

December 2023

* The authors thank Travis Box, Tony Cookson, Eugene McCarthy, Tavi Ronen (FMA Discussant), Matthew Serfling, Paul Tetlock, and seminar participants at the 2023 FMA Meeting, Auburn University, Clemson University, Mississippi State University, University of Tennessee, and Virginia Commonwealth University for helpful comments. All errors are our own.

HOW MANY WORDS IS A PICTURE WORTH?

USING EMOJIS FROM SOCIAL MEDIA TO PREDICT FUTURE STOCK RETURNS

Abstract

Using a new and comprehensive sample of more than 87 million Twitter posts referencing Russell 3000 firms between 2012 and 2022, we introduce a novel, unsupervised method of scoring the sentiment of emojis. Our method generates point-in-time dictionaries that map individual emojis to the contextual sentiment of recent tweets that contain them. In out-of-sample tests, we find that even controlling for the sentiment extracted from words, news, and corporate events, emoji sentiment correctly predicts future firm-level stock returns. Importantly, we show a newly emergent generation of Twitter users drive emoji-based return predictability, while more seasoned users better predict returns using words. Understanding the sentiment of emojis has become increasingly important as individuals and market professionals continue to adopt these new forms of communication.

Financial economists who study social media text and associated market outcomes must dredge through a rapidly evolving minefield of conventional words and newly adopted non-word tokens (*e.g.*, emojis) that contributors use to express their opinions and beliefs. For example, in the years 2021 and 2022, Twitter users collectively authored approximately 40 million original messages referencing a single Russell 3000 firm with a "cashtag."[1] Of these tweets, roughly 11.2 million contained words the standard Loughran and McDonald (2011) dictionary associates with either negative or positive sentiment. Interestingly, another 5.5 million tweets—which represent an incremental 49 percent of traffic—contained zero negative or positive words but included at least one emoji. And of the 11.2 million tweets employing scoreable words, 2.9 million also contained an emoji. With emoji use on the rise, most analyses of social media data, including both academic research and industry applications marketed to hedge fund clients, ignore these important tokens altogether.[2]

In this research, we assemble a comprehensive dataset of all tweets from 2012 through 2022 containing a cashtag reference to a Russell 3000 firm, and we introduce an unsupervised method for scoring the sentiment authors convey with emojis. Unlike "black box" mechanisms that incorporate proprietary machine-learning algorithms, our method is simple and transparent. We analogize our approach to a child learning to read by studying the context surrounding unfamiliar words. Thinking back to our preschool days, we recall encountering an unfamiliar word within a passage. Rather than seeking a dictionary-based definition, the young versions of ourselves would instead read surrounding context, which often comprised words that we

[1] In July 2023, Twitter was rebranded as "X." From that time forward, the company referred to platform messages as "posts" instead of "tweets." Since our sample ends prior to July 2023, we use the terms "Twitter" and "tweets" throughout this paper.
[2] The closest exception we can find is Hu et al (2021), who manually score the 300 most frequently used "tokens" in their sample of Reddit posts. These tokens can include both words and emojis.

understood, and use that understanding to extrapolate meaning to the new word. Our method of assigning sentiment to emojis follows the same pattern. When our text-parsing algorithm encounters an “undefined” emoji in the learning phase of the analysis, it infers the emoji’s sentiment from the text surrounding it. We collect inferred scores of all emojis during a specified time interval to create a vintage of our “point-in-time” emoji sentiment dictionaries. And since authors may alter their usage and intended meaning of various emojis over time, we repeat the “scoring” process each year in the sample to generate sequential vintages of the emojis sentiment scores.

Our emoji sentiment scores are sensible and intuitive. For example, our method assigns the moon emoji (🌕) a near-neutral score in the 2017 vintage but an increasingly positive score in each of the subsequent five annual vintages as social media authors embraced the phrase “to the moon” to convey optimism for a stock. Comparing commonly used pairs of emojis, we observe higher sentiment for a green light (🟢) than a red light (🔴), higher sentiment for a stock chart up (📈) than a stock chart down (📉), and higher sentiment for various iterations of a smiley face (😊) than a frowny face (☹️). More generally, we show that emoji sentiment aggregated across tweets at the firm-day level is positively correlated with both contemporaneous and past returns. Emoji sentiment is also positively correlated with *word* sentiment that we compute using two established sentiment dictionaries.

In our main analysis and as an example use case, we study the relation between emoji sentiment and future stock returns. Limiting the analysis to the years 2021 and 2022 due to coverage of a large cross-section of firms, we find that our novel emoji sentiment variable correctly anticipates future returns over the subsequent week. This effect is economically meaningful; a one standard deviation change of emoji sentiment corresponds to an incremental annualized return exceeding 5%. Extending the future return window through the next month, we find no evidence that this price movement reverses. Importantly, in all our return predictability tests, we employ a vintage of the emoji dictionaries that was created prior to the tweets used in the analysis. For example, we use the dictionary vintage corresponding to July 2020 through June 2021 when assigning emoji sentiment scores to tweets arriving from July 2021 through June 2022. This procedure separates the emoji scoring mechanism (which uses past data) from tests of whether emoji sentiment contains information about future returns. Equally important is the fact that emoji sentiment predicts returns irrespective of whether we control for the sentiment of the words within tweets, news sentiment, and corporate disclosure dates. Thus, emojis contain incremental information about future returns that methods which ignore an author's usage of these special characters currently discard.

In a second analysis, we group Twitter users into two cohorts depending on when they initially created their accounts and analyze which users' emoji use drives the return predictability result. Again limiting the analysis to the years 2021 and 2022, we show that emoji sentiment derived from users who created their accounts in 2020 ("New-Gen" users) correctly anticipates future returns, while emoji sentiment from users who created their accounts prior to 2020 ("Old-Gen" users) do not. Interestingly, we find the exact opposite pattern when we consider the sentiment of each groups' words or phrases. That is, the textual sentiment extracted from "Old-Gen" user Tweets correctly anticipates future returns while the textual sentiment from "New-Gen" user Tweets does not. We attribute these collective findings to each generation possessing unique—and evolving—abilities to communicate in their preferred dialect as opposed to one group having superior information.

We highlight our point-in-time dictionary methodology as a key contribution. While we score the sentiment of emojis in Twitter posts in the present application, the methodology is sufficiently nimble for alternative contexts. In its most general terms, the methodology requires three elements: (1) a list of tokens the researcher desires to score; (2) existing context that incorporates these tokens; and (3) a base mechanism that maps other contextual aspects to sentiment scores. In our data, the token list is a set of emojis; however, the list in other applications might include previously undefined words, words from another language, hashtags, non-character symbols, etc. While our context is a recent historical sample of Twitter posts, alternatives could include posts from alternative social media platforms, transcripts from recorded interviews, or any other text. Finally, our mapping mechanism is Hutto and Gilbert's (2014) VADER, which uses a static social sentiment dictionary created in 2014. Depending on the use case, one could instead map tokens to sentiment using a bag-of-words approach with a dictionary like Loughran and McDonald's (2011) sentiment lexicon or the Harvard Psychosocial Dictionary.

More broadly, our innovation relates to, and builds upon, other attempts to improve methods of sentiment extraction. For example, Garcia, Hu, and Rohrer (2023) devise a machine-learning approach to identify important positive and negative words used in earnings calls in association with contemporaneous stock returns. Their out-of-sample evidence indicates the new measure better explains price movements than an LM-based measure. Other researchers advance the literature by creating their own dictionaries that are more appropriate for a social media context. Two examples are Bradley et al. (2021) and Hu et al. (2021), both of whom study aspects of the Reddit forum Wallstreetbets.

Our emoji scoring methodology also appeals to regulators. For example, FINRA recently announced a new priority to monitor how member firms supervise off-channel communications between employees. Such communication includes the use of emojis to convey subtle messages that are not apparent in the written text. Executives in the broker-dealer space highlight the lack of an accepted emoji dictionary as a key challenge.[3] Likewise, recent high-profile lawsuits underscore the importance of emojis in investor communications and the need for an understanding of what such symbols convey. In 2022, billionaire investor Ryan Cohen tweeted a smiling moon emoji alongside his reference to Bed Bath and Beyond stock and then quietly exited his position in the company for an alleged $68 million profit. The court ruled Cohen and his fund must face a class action suit filed by other investors noting "Moon emojis are associated with the phrase 'to the moon,' which investors use to indicate 'that a stock will rise,' so meme stock investors conceivably understood Cohen's tweet to mean that Cohen was confident in Bed Bath and that he was encouraging them to act."[4]

A second significant contribution is the data itself, which represents an unprecedented message-level sample of how individuals communicate their views on the stock market via Twitter. Our entire dataset, which begins in 2012 when Twitter first introduced the use of cashtags, contains 87 million original tweets authored by more than 2.9 million unique Twitter accounts. The number of cashtag references to Russell 3000 firms has increased steadily from 12,000 per day in 2013 to over 82,000 per day in 2022. These statistics refer to original tweets only—the sample does not include retweets, which are essentially "forwards" of original tweets.

---

[3] Sun, Mengqi, "Wall Street Regulators' New Target: Emojis." *The Wall Street Journal,* June 29, 2023 (https://www.wsj.com/articles/emojis-wall-street-regulators-finance-finra-5bbf5688).

[4] Frankel, Alison, "Bed Bath and Beyond investor Ryan Cohen must face emoji-inspired shareholder suit." Reuters, July 28, 2023 (https://www.reuters.com/legal/litigation/column-bed-bath-beyond-investor-ryan-cohen-must-face-emoji-inspired-shareholder-2023-07-28/).

Moreover, while tweets are somewhat concentrated among a subset of firms, coverage is widespread. In 2013, Twitter users referenced about 1,200 unique firms per day. By 2018 and continuing through the end of our sample in 2022, coverage increased to 2,500 firms per day, or about 80% of the Russell 3000 universe. Emoji usage increases each year in the sample, as does the average number of firm days with a scorable emoji tweet. The sharpest increase in emoji usage occurs around the meme stock episodes of early 2021. Emojis only appeared in an average of 141,000 tweets per year between 2013 and 2019. This average increased to 500,000 in 2020 and over 3 million in 2021. Moreover, by the end of 2021, we observe emoji tweets for almost 2,000 *firms* per day.

Third, we offer new insights on the efficacy of two existing *textual* sentiment extraction methods within our Twitter sample. One method is a bag-of-words approach using the lexicons Loughran and McDonald (LM, 2011) constructed specifically for formal financial disclosure documents such as 10-K reports. The other is Hutto and Gilbert's (2014) VADER approach they developed specifically to analyze generic social media text. Both methods use publicly available dictionaries and open-source platforms. As such, our Twitter data analysis is particularly enlightening in the current environment where multiple data analytics firms use only proprietary algorithms to create their own social sentiment scores to market to industry clients.

We directly compare the utility of the two measures within the context of Twitter data by using them to predict *future* stock returns. For the full sample, the VADER sentiment score correctly anticipates future returns over the next day and week with no indication of a reversal through the first month. These results become economically stronger later in the sample as Twitter increased in popularity. Notably, the coefficient on LM's sentiment in similar regressions is significant only when predicting return over the first day. Thus, a simple VADER-based sentiment score emerges as a stronger and more robust predictor of future returns than a variable constructed using technology that targets formal text like firm 10-Ks.

Finally, we join a budding literature that examines how a variety of social platforms matter for financial markets. The Covid lockdowns of early 2020 saw a massive influx of first-time retail traders, who coincidentally faced lower-than-ever barriers to entry in wake of the brokerage industry's shift to zero-commission trading in the Fall of 2019. Anecdotally, we know small traders used social media to communicate and potentially coordinate trades during the meme stock episodes of early 2021. Some researchers have explored the influence of finance-focused platforms such as StockTwits, Seeking Alpha, and Reddit's Wallstreetbets (see Giannini et al. (2019); Chen et al. (2014); and Bradley et al. (2021), respectively). We have a far less developed understanding of Twitter, especially using tweet-level data. In early work, Chawla et al (2022) study tweets generated in 2013 and 2014 by major media outlets, accounts of S&P 1500 CEOs and CFOs, and those of S&P 500 companies themselves. They argue these tweets propagate stale news and contribute no new information into prices. In a more recent and comprehensive analysis, Cookson et al. (2022) utilize a proprietary black-box measure developed by industry covering the years 2012 through 2021. They find that Twitter sentiment predicts next-day returns, and the platform provides information that is distinct from that on StockTwits and Seeking Alpha.

I. Twitter Data

The sheer size and scope of our tweet-level dataset underscores our contribution.[5] Twitter allows academic users API access to its historical tweet-level data with a ten million query per month limit.[6] But as Twitter self-reports upwards of 500 million sent tweets per day as early as the year 2014, any researcher with reasonable time constraints must impose stringent filters prior to querying the data. Since we seek tweets related to individual stocks, we exploit authors' use of "cashtags" whereby they reference firms with a dollar sign followed by a ticker symbol. For example, the cashtag for Apple is $AAPL and that for Amazon is $AMZN. Twitter introduced cashtag references to its platform in July 2012, so we obtain a list of all Russell 3000 firms from the 2012 through 2022 constituent lists and query all tweets that include a reference to one of these firms.[7] We also only consider original tweets (i.e., no retweets) with available author data, and we require tweets be written in English. Finally, after the initial queries, we drop from our sample tweets that reference cashtags of more than one firm. The resulting dataset encompasses over 87 million original tweets authored by over 3 million unique user accounts.

Our initial queries return tweet-level and author-level variables. Tweet-level variables include the entire text, a timestamp, number of retweets, number of likes, an author identifier, attachments, and much more. Author-level variables include number of followers, number of tweets as of the query date, verified status, and the account's origination date. Due to Twitter's limit of 10 million records per month for each API user, our queries required several months of download time. The resources required to convert the raw JSON response from Twitter's API to workable data were similarly costly. To put this latter task into perspective, the processing

---

[5] Others using tweet-level datasets to analyze financial market outcomes include Gorodnichenko, Pham, and Talavera (2021), who study central bank communications, and Cookson et al (2023b), who study social media communication around bank runs.

[6] Twitter gradually discontinued this free access to academics in 2023.

[7] Newly reconstituted Russell indices take effect in late June of each year. Therefore, we use Russell constituent lists from June of year t for our queries of July of year t through June of year t+1.

pipeline running from the initial download to research-ready datasets consumed over 800GB of space. The resulting raw dataset is novel both in breadth – it contains an exhaustive record of all tweets mentioning a firm's ticker with the attached cashtag, and in timespan – it runs from the introduction of cashtags in 2012 through 2022.

In Figure 1, we plot in blue the total number of tweets referencing one Russell 3000 firm each day of our sample period. For the first eight years of our sample, cashtag references increased more than 17% per year from about 11,600 per day in 2013 to about 35,400 per day in 2020. The first visible spike in the figure coincides with the Gamestop episode of January 2021. On January 4th, there were 51,154 total tweets, and 98 of these referenced Gamestop. By the end of January, there were nearly 100,000 tweets per day, of which more than 14,000 contained "$GME". For all of 2021, Twitter activity averaged over 59,000 tweets per day, or a near 66% increase from the prior year. This number increased to more than 82,000 Tweets per day in 2022. Also in Figure 1, we plot in orange the number of firms per day with at least one tweet. In that series, we observe an increase from about 1,200 firms per day at the beginning of 2013 to around 2,500 firms per day, or over 80% of the Russell 3000, by 2018. Coverage remains at that saturation level through the end of our sample in 2022.

One advantage of our data is that we can see both traditional characters and emojis, the latter of which are represented as unique Unicode sequences.[8] Figure 2 summarizes the explosion of emoji usage over time. The blue line depicts the total number of tweets each day containing at least one emoji, while the orange line represents the number of firms with at least one emoji tweet. Prior to 2021, emoji usage was sparse. Between 2013 and 2020, about 237

[8] For example, the rocket ship emoji (🚀) appears in the data as "U+1F680". Emojis typically count as two characters toward Twitter's character limit of 140 characters (which increased to 280 characters in late 2017).

firms per day had tweets that contained emojis. Moreover, the number of Russell 3000 firms with an emoji tweet rarely exceeded 500 on any given day.

The Gamestop episode of January 2021 coincided with a radical shift. On January 4th, 5,711 tweets contained emojis. For perspective, on the average day between 2013 and 2019, only 559 tweets contained emojis. Even during 2020, emojis only appeared in 2,005 tweets on the average day. More interestingly, this shift is neither a temporary phenomenon nor is it limited to some small subset of meme stocks. Rather, it is a pervasive regime change, perhaps reflecting the new communication styles of individuals who entered the markets during the Covid shutdowns of the previous year. The orange line reveals the number of *firms* mentioned alongside an emoji nearly doubled in January 2021 to about 1,000 per day and continued to rise sharply to over 1,500 per day by the end of that year. To our knowledge, we are the first authors to document such changes in how people communicate about stocks on social media.

We next illustrate how the cross-section of firm-level Twitter activity has evolved over the past decade. To do this, we first collapse individual tweets to firm-day observations that represent the total number of tweets, the average tweet length (in words), the number of tweets containing an emoji (henceforth "emoji tweets"), and the average length of emoji tweets. For each quantity, we compute cross-sectional means, standard deviations, and various percentiles every day in the sample. Finally, we average each distributional statistic across days within three years: 2014, 2018, and 2022. We report the yearly averages of these distribution metrics in Table 1. We note these statistics are each conditional on a firm having a tweet (or a tweet containing an emoji) on a given day, so the average "n" in the tables reflect the growing sample sizes we observed in the orange plots from Figures 1 and 2.

Several notable patterns emerge. While many firms appear in at least one tweet on any given day, tweet intensity is skewed. The conditional mean number of tweets per firm increases from about 9 tweets per day in 2014 to about 28 tweets per day in 2022; the median increases during these same years from about 3 to 6 tweets per day. A similar skewness pattern exists for emoji tweets with means (medians) increasing from 2.5 (1.3) per day in 2014 to 10.8 (2.1) per day in 2022. Another salient result is the large number of firms experiencing only one emoji tweet per day, as the 25th percentile is roughly one tweet in each of the three years we report in the table. As Twitter restricted tweets to 140 characters until late 2017 and 280 characters thereafter, the tweets in our data are generally very brief messages. We report distributions of tweet length (in words) for each of the three years as well.[9] The average tweet length was less than 10 words in 2014 and 2018, but rose to about 14 words in 2022, attributable in part to Twitter's move that increased the character limit to 280. Medians behave very similarly to the means. Emoji usage is brief as well. In 2022, the average (median) emoji tweet contained only 2.5 (1.7) emojis and in all three years, the 25th percentile these for tweets slightly exceeds one emoji.

Researchers employ natural language processing algorithms to quantify a block of text's tone or sentiment. Popular methods rely upon established dictionaries containing individual word sentiment scores. For example, Hutto and Gilbert (2014) develop their Valence Aware Dictionary and sEntiment Reasoner (VADER) framework specifically for social media text. They created their dictionary using human raters who read and scored individual words within social media posts. As a second example, Loughran and McDonald (LM, 2011) provide a list of

[9] Single emojis typically count as one word in this calculus, but our methodology (described below) allows for blocks of emojis like diamond hands to show up as single words as well.

negative and positive words specific to finance contexts like firm 10Ks and other formal disclosures. Their word lists represent an evolution of the more general Harvard-IV Psychosocial Dictionary. Importantly, none of these dictionaries include emojis, so any application built upon these frameworks necessarily ignore beliefs and opinions social media authors convey using these powerful communication tools.[10]

We illustrate the extent of these omissions in Figure 3. Panel A provides a Venn diagram of the more than 40 million tweets in our data from 2021-2022. The orange circle represents about 26.6 million tweets containing at least one non-neutral word from the VADER dictionary. The blue circle represents tweets containing at least one emoji. While more tweets contain scorable words than emojis – 26.6 million vs 8.4 million – emojis are clearly an important communication tool. Moreover, about 2.3 million tweets in our sample contain emojis accompanied by no scoreable words.

Panel B paints a more striking picture with the commonly used LM dictionary. The orange circle represents tweets containing one or more word in either the LM negative word list (which consists of 2,345 words) or positive word list (which consists of 347 words). Thus, an algorithm utilizing the LM negative and positive word lists could assign a non-neutral sentiment score to each of these tweets. The blue circle, as before, represents tweets containing at least one emoji. Again we observe that more tweets contain scorable words than emojis at 12.2 million vs 8.4 million. Most important to our study, about 5.5 million tweets contain emojis without any

[10] We note that *emojis* differ from *emoticons*. Emoticons, which could be viewed as "ancestors" to emojis, consist of regular characters such as a semi-colon and a close parentheses ";)". Emojis are actual pictures, such as 😉. The VADER dictionary contains emoticons, but not emojis.

scoreable words. Thus, ignoring these tweets eliminates (5.5 / (5.5 + 12.2) =) 31% of potentially scoreable tweets from the data.

As the first large scale examination of Twitter data covering the cross-section of stocks, we conclude this section with a cursory analysis of the determinants of Twitter coverage. One plausible determinant is firm size, which is correlated with coverage by others such as analysts, the news media, and institutional investors. In Figure 4, we rank firms by market cap in December 2021 along the x-axis and then plot each firm's average number of tweets per day over 2022 using vertical bars. While one of the largest firms (Tesla) had the most tweets, and there is some decline in coverage moving left to right from the largest firms to the smallest firms, two other salient features emerge. First, many of the smallest firms still have substantial Twitter coverage. For example, AAR Corp, an aviation service provider that is near the middle of our size distribution averaged almost 150 tweets a day. Another example is Startek, an information technology company near the 5th size percentile that had 38 tweets per day (nearly10,000 tweets per year). In addition, firms in the bottom third of the size distribution had an average of over 13 tweets a day, which annualizes to over 3,300 tweets a year. Second, we can observe meaningful cross-firm variation in Twitter coverage that is independent of firm size.

II. Methodological Context

A. Overview of Point-in-Time Dictionaries

As Figure 2 depicts, emoji usage in social media has exploded of late, with the entry of a new generation of stock market participants during the Covid pandemic of 2020 and the meme stock episode of January 2021 potentially acting as catalysts. In 2020, Twitter users referenced 571 of our sample firms per day using emojis. That number rose to more than 1,400 firms per day in 2021 and exceeded 1,900 firms per day in 2022. A user might shift from a statement like

"I think $AAPL stock price is about to go up a lot" to the simply typing the cashtag "$AAPL" followed by nothing but a rocket ship emoji (🚀). Others might supplement normal text with one or more emojis within the same tweet. This practice is likely driven by both existing users shifting from words to emojis as well as the entrance of new users who simply communicate differently than incumbents. We formulate and test hypotheses along these lines in the Section IV.C below. Regardless, emoji use has quickly become commonplace in social media communication.

We propose a novel, unsupervised approach of assigning sentiment scores to the emojis and emoji sequences that appear in our Twitter sample. We liken our method to a child learning to read by using the context surrounding an unfamiliar word. Rather than seeking a dictionary-based definition, the child instead guesses the meaning of the new word based on the context that she already understands. For our purposes, the unfamiliar "words" are emojis, and the context is the voluminous Twitter text, whose sentiment we *can score* using an established mechanism such as VADER or LM. As we detail in Section III.A below, our method works as follows: We collect all scorable tweets containing emoji E_m during time period t, and we compute the average sentiment score S_{mt} for those tweets. We then assign the sentiment score S_{mt} to emoji E_m for time period t. We refer to the period-t mapping of sentiment scores to emojis ($S_{mt} \rightarrow E_{mt}$) as a *Point-in-Time Dictionary* since the mapping reflects the contextual sentiment of emojis over a particular time frame.

B. Measuring the Sentiment of Surrounding Text

Critical to our approach is the ability to gauge the sentiment of the text surrounding an emoji. That some emojis are used in tweets containing no scoreable text is not a problem so long as a sufficient mass of tweets contain both emojis and text. The Venn diagrams in Figure 3 indicate a substantial intersection of emoji and text tweets does in fact exist. Since there is no universally accepted established method of scoring social media text in a finance context, we consider two candidates: VADER and LM. Hutto and Gilbert (2014) created the VADER method to score social media text, and while commonplace in the more general NLP space, finance researchers have not yet considered its usefulness in contexts such as ours. In contrast, the finance literature has embraced Loughran and McDonald's (2011) method in its intended use of analyzing formal text such as 10Ks, but its efficacy in measuring sentiment in a social media context is unknown. Thus, which of these methods more appropriately scores tweets in our sample is an important empirical question. We now discuss the VADER and LM methods in greater detail and evaluate their usefulness in our Twitter sample.

1. *Valence Aware Dictionary and eSentiment Reasoner (VADER)*

Hutto and Gilbert's (2014) Valence Aware Dictionary and sEntiment Reasoner (VADER) framework utilizes a dictionary of negative and positive terms. The authors developed their dictionary through a crowdsourcing marketplace called Amazon Mechanical Turk whereby human raters read and scored individual words within social media text according to their perceived tone. With this dictionary in hand, the VADER algorithm reads text and computes a raw score according to the words therein. It then transforms the score according to five key heuristics of microblogging and squeezes the output via a nonlinear transformation into a number between -1 and 1.[11] The VADER dictionary is easily accessible to any researcher wishing to

[11] The VADER model takes into consideration 5 heuristics prior to passing the intensity scores through the non-linear function. These heuristics are: punctuation, capitalization, degree modifiers, shifts in polarity from "but", and shifts in polarity from trigrams.

score text. Moreover, we invoke its algorithms and heuristics to score all tweets via open-source Python packages.

We collapse the data to the firm-day level as follows. First, we define the variable $VSent\_Tweet_{ij,t}$ as the VADER-based sentiment score for tweet j that references firm i on day t, and it ranges from -1 (extremely negative) to +1 (extremely positive) with a score of zero representing neutral sentiment. Then, for each firm i and day t, we average all tweet-level sentiment scores to compute the variable $VSent_{\mathrm{it}}$, which also ranges from -1 to 1. We present average cross-sectional distribution characteristics of *VSent* for the years 2014, 2018, and 2022 in Table I Panel B. These cross-sectional distributions are conditional on a firm having a *scorable* tweet on a given day. Thus, for *VSent*, we require at least one tweet containing a word from VADER's 7,500-word dictionary. We can see again that coverage at the firm-day level increases modestly throughout our sample. The VADER method scored tweets for 1268 firms, 2,195 firms, and 2,338 firms on an average day in 2014, 2018, and 2022, respectively. The fact that these numbers are noticeably lower than the number of firms having tweets indicates the VADER dictionary imposes modest constraints on our sample.

One salient observation regarding *VSent* is that the distribution is tilted toward positive sentiment. With a potential range of -1 to +1 and zero representing neutral sentiment, the mean and median values for *VSent* are in the vicinity of 0.2 to 0.3. Moreover, zero typically falls between the 5^{th} and 25^{th} percentiles. Comparing the statistics across the three years, the cross-sectional distribution of *VSent* appears quite stable throughout our sample. The positive tilt of Twitter sentiment is consistent with studies of other social media datasets including Yahoo! Finance, Raging Bull, StockTwits, and Seeking Alpha (e.g., Antweiler and Frank, 2004; Cookson and Niessner, 2020; Hu et al., 2021). Notably, positive average social sentiment contrasts with our understanding of financial news sentiment (e.g., Tetlock, Saar-Tsechansky, and Macskassy, 2008).

2. *Loughran and McDonald (2011)*

Loughran and McDonald (2011) also offer a dictionary-based approach. Specifically, they identify words that appear in at least 5% of a large sample of Firm 10-Ks and then, upon inspection, they flag those they deem to have negative connotations within a finance context. This list is similar in spirit to the Harvard-IV-4 TagNeg file other researchers use in more general contexts. Researchers commonly use the LM lexicons to measure a block of text's sentiment in a bag-of-words fashion. Specifically, one could define the "negativity" of a passage as the fraction of its words appearing in the LM negative word list. In their original paper, LM also consider the fraction of positive words, but they argue that variation in negativity captures the most relevant information. Others focus only on the negative word list as well (see, e.g., Chen, De, Hu, and Hwang's, 2014, analysis of Seeking Alpha posts). Tetlock (2007) argues a similar point using the more general Harvard-IV dictionary.

We compute the variable *LMNeg_Tweet* as for each tweet in our data as the fraction of negative words from the LM negative word list and average across tweets within a firm day to calculate *LMNeg*. We include distributional statistics for *LMNeg* in Table I Panel B. While the LM dictionary contains over 86,000 words, all but 2,692 of these words are neutral, and hence, the *LMNeg* variable extracts no sentiment for many firm days. For example, in 2014 almost half of the firm days scored by *LMNeg* have no tweets that contain any non-neutral words; almost half the cross-section has a score of zero even in 2022. In looking back at Figure 3, we observe a similar pattern at the tweet level during 2021 and 2022.

C. Correlations

We assess the viability of *VSent* and *LMNeg* to capture social media sentiment through their correlation with contemporaneous stock returns (ret_0), and stock returns from the prior day (ret_{-1}), week ($ret_{-5:-1}$), and month ($ret_{-26:-6}$). Recognizing that Twitter sentiment could both lead and lag price movements within the same day, we use the term "contemporaneous" quite loosely. To mitigate the effects of noise, we impose two additional filters to the creation of our sentiment variables for this analysis and the analyses that follow. First, we remove tweets generated by accounts that posted more than one hundred times per day in the prior calendar month. We use this filter to avoid contaminating our measures with tweets created by bots. Second, after collapsing tweets to the firm-day level, we only keep firm days having at least two tweets. We employ this latter filter to avoid giving too much emphasis to any single author.

In Table 2, we report daily cross-sectional correlations between both sentiment variables and stock returns averaged within each year of our sample. As with the distributions in Table 1, these correlations are conditional on the existence of Twitter activity for the firm-day. We observe the correlations between *VSent* and the return variables are uniformly positive and stable over time. Likewise, those for *LMSent* are uniformly negative. Thus, the variables reflect information within both contemporaneous returns and past returns with the signs one would expect for sentiment variables. Interestingly, the magnitudes of correlations with contemporaneous, prior day, and prior week returns are roughly the same as one another. And the correlations with prior month return ($ret_{-26:-6}$) are only marginally smaller. Thus, the correlations are consistent with social media commentary that reflects events of the recent past.

We also report correlations between *VSent* and *LMNeg* in Table 2. Average correlations range from -0.24 in 2017 to -0.33 in 2015. These correlations indicate *VSent* and *LMNeg* contain common information; very low values of *VSent*, which represent negative tone, are associated with the usage of a high fraction of negative words, i.e., high *LMNeg*. Since LM provide a list of positive words, we compute two additional variables: *LMPos*, as the fraction of words used that are positive, and *LMTone*, as the difference between *LMPos* and *LMNeg*. The final two columns of Table 2 report correlations with each of these measures. These correlations are consistently positive, again indicating that *VSent* and the LM bag-of-words measures contain some common information.

Moving forward, the correlations between *VSent* and the LM measures prompts more questions. While one can argue a VADER-based measure is more appropriate for gauging the sentiment of social media text, which measure performs better in empirical applications? In the next section, we utilize these variables to explore whether Twitter sentiment forecasts *future* stock returns, and in doing so we compare the usefulness of *VSent* and *LMNeg* in a return predictability model. This exercise captures perhaps the most basic question researchers and industry practitioners ask with respect to social media and the financial markets. For our immediate purposes, however, the exercise informs on which measure is potentially more useful for our task of scoring emojis.

D. Textual Sentiment and Future Stock Returns

We next analyze the relationship between Twitter sentiment and future cumulative abnormal stock returns (*CAR*) by estimating the following panel regression:

$$CAR_{i,t+k1:t+k2} = Sentiment\_Variable_{i,t} + Controls_{i,t} + \varepsilon_{i,t} \tag{1}$$

Our main sentiment variables are $VSent_{it}$ and $LMNeg_{it}$, though in unreported results, we also consider the variable $LMTone_{it}$, which captures variation in both the negative and positive word lists from LM. The dependent variable $CAR_{i,t+k1:t+k2}$ is the sum of daily CRSP value-weighted market-adjusted returns over the interval (t+$k1$, t+$k2$). Our primary controls are returns over several horizons: ret_{it}, ret_{t-1}, $ret_{t-5:t-1}$, $ret_{it-26:t-6}$; market equity as of the most recent December: ln(ME_{it}); and book-to-market equity using the prior fiscal year end book value and prior June market equity: ln(BM_{it}).[12] We include day fixed effects and compute Newey-West (1987) standard errors with leads and lags selected to match the horizon of our dependent variable.

We report point estimates from Equation (1) in Table III. In Panel A, we relate day t Twitter sentiment to future returns over the next day (CAR_{t+1}) and the remainder of the first week ($CAR_{t+2:t+5}$) using the full sample from 2013 through 2022. The estimates indicate Twitter sentiment correctly anticipates returns over the very short-term as the coefficient on $VSent_t$ in the first column is a positive and statistically significant 5.1. Importantly, we control for the positive correlation between sentiment and both contemporaneous and prior period returns. Insofar as intraday price movements on day t occur subsequent to same-day tweets, our inclusion of day t return coupled with the analysis of a future return staring on day t+1 understates the ability of Twitter sentiment to predict future returns. In column (4), we observe the ability of *VSent* to predict future returns persists through at least the first full week as its coefficient remains positive and statistically significant (15.7) when the dependent variable is $CAR_{t+2:t+5}$. To gauge the economic significance of this predictability, we multiply the coefficient in column (4) by the standard deviation of *VSent* (0.24) and then annualize the product. This results in an incremental cumulative abnormal return of 15.7 x 0.24 x (252/4) = 2.4% annualized.

[12] We obtain return and market cap variables from CRSP and book value of equity from Compustat.

How this *VADER*-based sentiment metric performs compared to an alternative measure constructed using the LM negative word list is an important empirical question whose answer is particularly relevant to researchers designing future studies of social media sentiment and the stock market. We therefore re-estimate Equation (1), but we replace *VSent* with *LMNeg* and report the estimates in columns (2) and (5). Compared with that of *VSent*, the ability of *LMNeg* to predict future returns is short-lived. The coefficient on *LMNeg* becomes only marginally statistically significant after the first day. Moreover, when we include both *VSent* and *LMNeg* in the same model, the coefficient for *VSent* remains statistically significant when predicting both CAR_{t+1} and $CAR_{t+2:t+5}$ and actually increases slightly in magnitude. In contrast, the coefficient on *LMNeg* weakens in magnitude though it does retain significance when predicting next-day returns.

The facts that Twitter activity containing cashtags has intensified through time (see Figure 1) and emoji-usage has grown from almost non-existent in 2013 to commonplace in 2021 and 2022 (see Figure 2) motivate us to study how the patterns depicted in Panel A have changed during our sample. We therefore re-estimate the model from Column (4) in Table III Panel A each year separately. We plot the coefficient estimates for *VSent* and their 95% confidence bands in Figure 5. We see from the figure the relation between VSent and future returns does tend to strengthen in the latter years, though the relationship is somewhat volatile from one year to the next and the trend is not monotonic over time.

Our findings that *VSent* and to a lesser extent *LMNeg* predict the cross-section of future returns are interesting, but not entirely novel. In related work, Cookson et al (2023) show that sentiment from three large platforms – StockTwits, Seeking Alpha, and Twitter – each contains unique information about future stock returns. However, those authors rely upon a proprietary measure from the data analytics firm Social Market Analytics to quantity Twitter sentiment. Similarly, Gu and Kurov (2020) show that Bloomberg's machine-learning based measure of Twitter sentiment correctly anticipates stock returns. In contrast, our return predictability analysis utilizes completely open-source methods (i.e., VADER and LM) to extract Twitter sentiment.[13] Thus our analysis is both instructive and encouraging for future researchers and industry participants who wish to develop their own algorithms for measuring Twitter sentiment.

Of course, we do not intend our text-based return predictability analysis to be the final say on how to best use Twitter data to predict stock returns or even on how to measure the association. And at this point, we remain somewhat agnostic on the economic interpretation of those relationships. Rather, we simply use these findings to establish the VADER-based *VSent* score as a viable ingredient in building our point-in-time emoji dictionaries described in detail below. We trust that future work, possibly our own, will improve upon the extraction of value-relevant sentiment from social media data and better illuminate the underlying economics. Moreover, since our dictionary approach is both general and adaptable, any future improvements in text-based sentiment scoring will undoubtedly enhance our method's practical usefulness.

III. Point-in-Time Emoji Dictionaries

A. Dictionary Summaries

Turning to our main objective, we use the Twitter sample to create point-in-time emoji dictionaries. While we outlined our dictionary approach in general terms above, our use of

[13] Cookson et al 2023b separately demonstrates the usefulness of the VADER methodology in predicting future stock returns.

VADER-based sentiment scoring in this implementation warrants a more specific restatement of our methodology. The steps are as follows:

1) Each year *t*, we identify via Unicode signatures every emoji *m* that appears in the Twitter data. Then, we collect the VADER scores of all Tweets containing a given emoji. We refer to these tweets as "context tweets."
2) For each emoji *m* and context tweet *w*, we pass the context tweet's VADER score through the inverse of VADER's non-linear function. This step generates the sum of the individual word scores in that tweet after accounting for all the heuristics that the VADER model uses. We refer to this outcome as the tweet's raw VADER score.
3) Again for each emoji *m* and tweet *w*, we divide the raw VADER score from step 2 by the number of scorable words in the context tweet, which produces the average intensity for the words that surrounds emoji in the context tweet. We refer to this outcome as the tweet's VADER word intensity.
4) For each emoji *m*, we compute the mean of the VADER word intensities across all context tweets. The resulting quantity becomes that emoji's sentiment for year *t*.

The final sentiment score for emoji *m* in year *t* (E_{mt}), therefore, is essentially an "average" sentiment of the words surrounding the emoji in that year. The collection of emoji sentiment scores for a given year represents our point-in-time dictionary for that particular year as the scores reflect emojis' contextually-derived sentiment at a specific point in time. In subsequent analyses, to avoid any look-ahead bias in our analysis, we use sentiment scores from year *t* point-in-time dictionaries when relating the sentiment of emojis in year *t*+1 to financial market outcomes.

Before proceeding, we should clarify three details about our dictionary construction. First, context tweets can (and often do) contain more than one emoji. When this occurs, we use a single context tweet multiple times in the scoring process – once for each emoji present. Second, we separately consider "emoji blocks," which we define as multiple emojis of at least two types that appear in a single uninterrupted string. That is, we parse all text to identify emoji blocks that authors use regularly, and we assign each block its own sentiment score. For example, tweets often contain a diamond followed by hands (💎🙌). A user might include the "diamond hands" to indicate the desire to hold on to a position. The exception is an uninterrupted string of the same emoji with no other emojis (e.g., 🔥🔥🔥). We do not separately score such sequences. Third, when creating the dictionary for year t, we obtain context tweets from July of year t -1 through June of year t. While this choice is arbitrary, we selected July through June to coincide with a single Russell 3000 vintage.

We also note the length of the time interval containing the context tweets and the frequency with which we update the dictionaries are both choice parameters for the estimation. When selecting these parameters, we considered the following tradeoff. On the one hand, our desire for precise sentiment scores for a large sample of emojis warrants a relatively long window for context tweets. For example, in the 2022 vintage of our emoji dictionary included 1,481emojis, or emoji blocks, using more than 11.2 million occurrences within scorable context. Shortening the context window limits the number of emojis we can reliably score. On the other hand, we recognize emoji sentiment likely changes at a much higher velocity than the sentiment authors associate with traditional words. This potential evolution motivates shorter context windows with more frequent updates. We view our choice of one-year context windows with annual updates as a middle ground that balances the tradeoff. We have also experimented with three-month rolling context windows (i.e., use January through March context tweets for the April vintage, February through April context tweets for the May vintage, etc). For comparison to our annual vintage, the June 2022 three-month dictionary included only 664 emojis, or emoji blocks, with 3.4 million occurrences. From a use-case perspective, these dictionaries are also viable; they produce similar inferences as the annual vintages used in Table 6 below.

We summarize our point-in-time dictionaries in Table IV. For each year, we include the total number of emojis and emoji blocks scored with at least 100 tweets. We also report the total number of tweets used (to score all emojis and emoji blocks) and the average sentiment score, again recognizing that an individual tweet may be used multiple times to score each emoji therein. We note that the average sentiment score is positive, which reflects the underlying tilt toward positive tweets (see Table 2). Just as the number of tweets containing emojis and the number of firms referenced by emoji tweets explode in the final two years of the sample, we see in Table 4 that the number of unique emojis appearing in at least 100 tweets rises dramatically near the end of the sample period. Between the 2020 and 2021 vintages of our emoji dictionaries, the number of unique emojis scored almost doubles from 396 in 2020 to 675 in 2021. This number continues to increase to 886 in the final vintage (2022), which draws from almost 11 million tweets between July 2021 and June 2022. The number of unique emoji blocks also reaches its maximum in the final vintage with 595 scored blocks appearing in a total of almost 300,000 tweets.

B. Emoji Sentiment Examples

Prior to any statistical analyses, we provide a few anecdotal examples of the scores our method assigns to specific popular emojis.

Example 1: The rocket ship (🚀) was the most common emoji in each of the final two vintages with almost 500,000 uses in 2021 and nearly 1.6 million uses in 2022. Its score is 1.35 and 1.44 in these two dictionaries, respectively.

Example 2: The moon (🌕) became popular in the last two vintages as authors used the phrase "to the moon" as a way of showing extreme optimism for a stock. In 2022, the moon was used over 20,000 times and had a score of 2.30. Its usage and score represented marked increases from 2,447 tweets and 2.03 just one year prior.

Example 3: The stock chart up (📈) and stock chart down (📉) are very clear indicators of beliefs about future price movements. In 2022, these emojis were each used in more than 180,000 tweets, and their relative scores of 0.40 and -0.53 appropriately reflect their opposite directions.

Example 4: The green light (🟢) and red light (🔴) were similarly popular in 2022 with appearances in about 38,000 and 51,000 tweets, respectively. Their scores in that dictionary are 1.09 and -0.90.

Example 5: The smiling face (😊) and angry swearing face (🤬) are two examples of many versions of "happy" and "sad" faces. These appear in about 12,000 and 2,000 tweets in 2022 with scores of 1.41 and -0.66.

C. Correlations

So how useful are our point-in-time emoji sentiment dictionaries? In this section, we create firm-day sentiment scores and correlate them with other variables we expect to capture a dimension of sentiment. Then, in Section IV below, we consider whether emoji sentiment predicts future returns. We begin by computing tweet-level emoji sentiment, $ESent\_Tweet_{ijt}$, following the standard VADER algorithm. Specifically, for a given tweet, we obtain individual emoji sentiment scores from the relevant point-in-time dictionary. To circumvent the effects of year-to-year fluctuations in the tone of context tweets, we demean each emoji's sentiment by the average sentiment from its point-in-time dictionary. We then modify scores according to VADER's microblogging heuristics and pass the sum each tweet's scores through VADER's non-linear function to output a tweet-level score between -1 and 1. Finally, for each firm *i* and day *t*, we average all $ESent\_Tweet_{ijt}$ to compute the firm-day emoji sentiment, $ESent_{it}$, which also ranges from -1 to 1. We emphasize that the variable $ESent_{it}$ measures the sentiment *only of the emojis* appearing in tweets about firm *i* on day *t*. It ignores the sentiment of other words tweet authors use. Our original variable $VSent_{it}$ captures the sentiment of those words. As previously, we eliminate tweets from accounts that averaged more than 100 tweets per day over the prior month, and we require at least two scored tweets to retain the *ESent* measure for a given firm-day.

We report in Table 5 average daily cross-sectional correlations between our new variable *Esent* and contemporaneous and past returns as well as the sentiment of the actual words in the tweets, *Vsent*, *LMNeg*, *LMPos*, and *LMTone*. Similar to *VSent*, our new variable *ESent* is positively correlated with same-day returns and prior week returns. More importantly, we report that *ESent*, which only measures the sentiment of emojis, and *VSent*, which only measures the sentiment of words exhibit positive contemporaneous correlations ranging from a low value of 0.09 in 2015 to a high value of 0.26 in 2017. Thus, the word-based and emoji-based sentiment measures contain some common information. Finally, *ESent* and the LM sentiment variables capture some common information, though the correlations between *ESent* and the LM variables are notably smaller in magnitude than the correlations between *ESent* and *VSent*.

IV. Do Emojis Predict Future Stock Returns?

As a practical use case, we now analyze the relationship between *ESent* and future stock returns. This relationship is interesting for multiple reasons. First, prior research indicates the informational content in social signals is multidimensional. Just as Cookson et al (2023) argue that users posting on different social media platforms draw from unique information sets, individuals who communicate using emojis could uncover information that is orthogonal to that revealed by others who communicate using words. Second, the return predictability analysis taps into a broader question about whether the evolution of language to include new summary "pictures" is beneficial in terms of communicating insights or alternatively washes out important details that are more cleanly stated with words and sentences. Third, the direction of the relationship between *ESent* is an empirical question. While emojis could convey meaningful new information about fundamental values, the readers of tweets might also respond in a primitive and emotional way to an emoji and trade in a manner that drives prices away from fundamentals.

A. Regressions

Since emoji usage on Twitter does not span a reasonably large cross-section of firms on a typical day until 2021, we estimate Equation (1) for the 2021-2022 time period using *ESent* as a sentiment variable. We also estimate models including either *VSent* or *LMNeg* to better highlight the information emojis convey that is incremental to what we can already extract from words. We report the results for various horizons of future returns in Table 6. The salient message of this table is that emoji sentiment, as measured by our unsupervised learning approach, correctly anticipates future stock returns over the subsequent week. We observe in Panel A that both when CAR_{t+1} is the dependent variable (the first three columns) and when $CAR_{t+2:t+5}$ is the dependent variable (the final three columns), the coefficient on *ESent* is positive and statistically significant. This result holds when *ESent* is the only sentiment variable in the model and when we include variables that capture the sentiment of words (*VSent* and *LMNeg*) as controls. The latter result is particularly meaningful because it suggests *ESent* conveys information about future stock returns that is incremental to the information we can extract from the words used by tweet authors.

The economic magnitudes of our coefficient estimates are impressive. For example, the *ESent* coefficient in Column (5), which is the model controlling for *VSent* along with other standard controls, is 46.1. A one standard deviation change in *ESent* therefore corresponds to an incremental CAR of 46.1 x 0.15 x (252/4) = 4.4% annualized.[14] To our knowledge, this is the first-ever reported large-sample evidence that emoji sentiment predicts stock returns. We expand this analysis in Section IV.C below with a preliminary exploration of which Twitter users contribute to the emoji-based return predictability.

At least two interpretations fit our result that Twitter emoji sentiment correctly anticipates stock returns. On the one hand, aggregating the very noisy beliefs of the individuals who write Twitter posts can generate an informative signal about stock prices, and in that sense, the sentiment captured by *ESent* is "correct." On the other hand, irrational traders could respond to the sight of emojis and trade in the same direction as their implied sentiment to the point that prices move past intrinsic values. The former would be consistent with recent research indicating

[14] The economic magnitude of the *VSent* coefficient in this same model is similarly impressive. A one standard deviation change in *VSent* corresponds to an incremental CAR of 37.0 x 0.24 x (252/4) = 5.6% annualized

that retail investor trading imbalance correctly anticipates future returns (e.g., Kelley and Tetlock, 2013; Boehmer, Jones, Zhang, and Zhang, 2021), while the latter would represent a more general case exemplified by the recent meme stock episodes in which traders bid up the prices of a few stocks in response to comments on Reddit's Wallstreetbets. Differentiating between these interpretations is critical to our understanding of the relationship between Twitter sentiment and financial market outcomes as only the former suggests social media improves the incorporation of information into prices.

If the positive coefficients on emoji sentiment in Table 6 Panel A result from irrational traders moving prices in response to emojis, the resulting price movements should be reversed over some future horizon. That is, there should be some return horizon after the first week for which the sign on the sentiment coefficient is negative. To help distinguish the two interpretations, therefore, we extend the return window to cover the full month after we observe emoji sentiment. In Table 6 Panel B, we report results from the same models as in Panel A, except with $CAR_{t+6:t+10}$ and $CAR_{t+11:t+20}$ as the dependent variables. Importantly, the coefficient on *ESent* is insignificant after the first week. While still preliminary in nature, the upshot from these estimations is that across the various time horizons and specifications, the relation between emoji sentiment and future returns does not reverse.

To better depict the return predictability dynamics, we estimate the model separately using $CAR_{t+1:t+k}$ for each k = 1 through 20 as the dependent variable and save the coefficients for *ESent*. We plot the series of coefficient estimates along with 95% confidence bands in Figure 6. There, the coefficient corresponding to the k=1 value on x-axis is the estimate for *ESent* when CAR_{t+1} is the dependent variable; that for the k=2 value is the estimate when $CAR_{t+1:t+2}$ is the dependent variable; and so on. The figure illuminates two clear patterns. First, the coefficients steadily increase for about the first 10 days. This pattern indicates that *ESent* has contains similar information about each of subsequent 10 days. Second , after day ten, the coefficient remain roughly constant. This pattern suggests *ESent* has little information about returns after the second week and that the stock price movements in the first 10 days are permanent through at least a month after we observe *ESent*. Together, these results reinforce our initial interpretation that Twitter emoji sentiment contains useful information about future stock returns.

B. Alternative Information Releases

One plausible alternative explanation for our results in Table 6 is that Twitter users observe some other visible signals and post emojis to summarize already publicly available information. Some examples of this information are news stories and firm disclosures. Such a result would still be interesting in the sense that Twitter users relay information through their social networks, but it would alter our interpretation that Twitter users convey novel information through emojis. We consider this mechanism by controlling for three alternative sources of information releases: news stories, firm-level 8k disclosures, and earnings announcements.

We obtain Dow Jones Newswire dates and sentiment scores from Ravenpack and dates for 8K filings and earnings releases from EDGAR and we augment our main model accordingly. The variables $DJNWNum\_Stories_{it}$ and $DJNWSent_{it}$ represent the number of individual stories and their average sentiment for the day respectively, and $DJNWSent_{it}$ equals zero if no stories appear on day t. The variables $EADate_{it}$ and $8\text{-}KDate_{it}$ are indicators set to one if day t contains an earnings release or 8-K, respectively, and zero otherwise. We interact each of these indicator variables with ret_{it} with dummy variables as a proxy for announcement's information content.

We present the results in Table 7 using CARs with the same horizons as before as the dependent variable. Consistent with prior literature (e.g., Tetlock et al, 2008), the tone of news stories correctly anticipates future returns over short horizons. Similarly, the interaction coefficient estimate for $8\text{-}KDate_{it}$ x ret_t is significantly positive when predicting next-week returns, indicating a gradual incorporation of the information content of firm announcements. In contrast, we find no relation between the tone of earnings announcements and near-term returns. The most important findings in Table 7 are the *ESent* coefficient estimates. When predicting both CAR_{it+1} and $CAR_{it+2:t+5}$, the estimate is positive and statistically significant, and of a magnitude similar to the estimates from Table 6. Together, these results suggest emoji sentiment correctly predicts next week returns, and this predictability is incremental to the information from other releases from the press and firms.

C. Which Users Matter?

The post-2020 insurgence of emoji usage and its striking ability to predict the direction of future stock returns command additional analysis. In this final section, we explore two non-mutually exclusive hypotheses that can explain these interesting patterns in the informational content of social media communication. On the one hand, incumbent Twitter users' inclusion of emoji usage could convey unique value-relevant information they would not otherwise reveal using words. This mechanism is largely statistical in nature to the extent that our analysis of emojis removes measurement error inherent in a strict analysis of text. On the other hand, an entirely new generation of individuals could enter the fold and bring with them a fresh vocabulary that includes emojis. By communicating differently than their predecessors do, this new generation might convey unique information through emojis. This is an economic mechanism that links the information within emojis to a certain type of individual.

We separate Twitter users into generations according to when, relative to the year 2020, they created their accounts.[15] We label users who created their accounts prior to 2020 "Old-Gen" users and those who created their accounts during 2020 "New-Gen" users. To hold each group's composition constant during the analysis period of 2021 to 2022 below, we ignore all accounts created after December 31, 2020 for these tests.[16] We define user groups in 2020 for three reasons. First, the Covid pandemic, which began in the US in early 2020, represented a time when a new generation of individuals entered the financial markets, due in part to the shutdowns that forced many to remain at home for much of the year. Second, this period immediately preceded the rampant rise of emoji usage that we observe in Figure 2. Third, and more practically, separating users into cohorts based on information in 2020 allows us two full years of data to compare the groups to one another.

In Figure 7, we compare Old-Gen and New-Gen users' propensities to incorporate emojis in their tweets. Panel A displays each group's time series of emoji tweets, scaled by the number of users in the respective group. We observe that New-Gen users (depicted by the orange line) produce far more emoji tweets per user than the Old-Gen users (depicted by the blue line) produce. Over the two years, New-Gen users average 7.1 emoji tweets a day per 1,000 users while Old-Gen users average only 2.7 emoji tweets a day per 1,000 users. Panel B reveals these patterns are not driven simply by New-Gen users tweeting more. On an average day in the sample, New-Gen users include emojis in 23.3% of their tweets, while Old-Gen users only include emojis in 17.8% of their tweets. Our main takeaway from Figure 7 is that while both

[15] We obtain the date each user account was created via Twitter's API.
[16] Our results are not dependent upon this choice. When we also include users who created their accounts after December 31, 2020 in the "New-Gen" group, the results described below are generally stronger.

Old-Gen and New-Gen users incorporate emojis in their tweets, New-Gen users do so to a much greater degree.

We recompute our *ESent* and *VSent* variables separately for each group and repeat our main return predictability analysis in Table 8. Models appearing in the odd-numbered columns contain Old-Gen *ESent* and *VSent* variables; those appearing in the even-numbered columns contain New-Gen *ESent* and *VSent* variables. Two salient results emerge. First, like the main result from Table 6, the coefficient estimate for *ESent*(New-Gen) is positive and statistically significant when predicting returns for the first week. In stark contrast, the same coefficient estimate for the Old-Gen users is statistically (and economically) zero. Thus, these results suggest that the newer adopters of Twitter drive the relation between emoji sentiment and future stock returns. This finding is interesting and useful as it suggests a new form of communication, which is brought forth by new financial market participants, is informative.

We illustrate the difference between emoji-based return predictability across the two groups by once again estimating the model separately using $CAR_{t+1:t+k}$ for $k = 1$ through 20 as the dependent variable and save the coefficients for *ESent*. We plot in Figure 8 the series of New-Gen coefficient estimates in orange and the series of Old-Gen coefficient estimates in blue. Differences across groups are stark. The patterns for the New-Gen coefficients resemble those from Figure 6 that represent *ESent* constructed from the full sample of Twitter users, albeit the magnitudes are larger. In contrast Old-Gen *ESent* contains no information about future returns throughout the subsequent four weeks.

Does the strong predictability of New-Gen *ESent* arise from these users' superior information processing ability in general or is it specific to their communication through emojis? The second salient result in Table 8 addresses this question. Turning to the coefficient estimates for *VSent*, or the sentiment conveyed through words, we observe a very different pattern of that in *ESent*. The coefficients for *VSent*(*Old-Gen*) when predicting CAR_{t+1} and $CAR_{t+2:t+5}$ are 10.7 and 38.9 with *t*-statistics of 1.63 and 2.36, respectively. In contrast, the coefficients for *VSent(New-Gen)* are never statistically significant. Thus, while New-Gen users convey superior information through their use of emojis, Old-Gen users convey superior information via words.[17] In sum, the two cohorts of Twitter users do not differ in whether they produce value-relevant information in their tweeting activity. Rather, they differ in which form of communication – words for the Old-Gen versus emojis for the New-Gen – contains value-relevant information.

V. Conclusion

Social media has revolutionized how people communicate. The transformation that began with a shortening of formal speech and the elimination (or excessive use) of punctuation and capitalization led to the creation of new forms of slang and the complete redefinition of established words. And most recently, the emergence of emojis eliminates some usage of definable terms altogether. Clearly, our understanding of how this type of communication reflects, predicts, and potentially even determines financial market outcomes is married to our ability to objectively map an evolving language into measurable quantities. The emoji-scoring approach developed in this paper makes initial progress.

We view this paper as but the tip of the iceberg. Our point-in-time dictionaries that link emojis to sentiment scores according to their contextual usage, and the initial results that these

[17] We repeat this exercise only using individual tweets that contain both scoreable words and emojis. While drastically reducing the number of tweets in the analysis, the constrained sample arguably offers a more precise test of specific individuals' usage of words versus emojis. In this analysis, we find economically stronger results that those we report in Table 8.

sentiment scores predict future returns in out-of-sample tests prompt many future questions. What do social media authors desire to convey with pictures that they cannot or will not convey with words? Do Twitter users respond differently to viewing a picture than they do to reading text? And given the rapid evolution of slang, can one use our same method to effectively extract sentiment from new words or phrases or even track changes in the connotations of well-established language? We look forward to participating in the future exploration of questions such as these.

References

Antweiler, W., and M. Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, Journal of Finance 59: 1259-1294.

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

Bellstam, G., S. Bhagat, and A.J. Cookson, 2021, A text-based analysis of corporate innovation , Management Science 67(7): 3982-4642.

Boehmer, E., C.M. Jones, X. Zhang, and X. Zhang, 2021, Tracking retail activity, Journal of Finance 76(5): 2249-2305.

Bradley, D., J. Hanousek, R. Jame, and Z. Xiao, Place your bets? The market consequences of investment research on Reddit's Wallstreetbets, University of South Florida working paper.

Chawla, N., Z. Da, J. Xu, and M. Ye, 2022, Information diffusion on social media: Does it affect trading, returns, and liquidity? University of Notre Dame working paper.

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

Cookson, J. A., Fox, C., Gil-Bazo, J., Imbet, J. F., & Schiller, C. (2023). Social media as a bank run catalyst. *Available at SSRN 4422754*.

Cookson, J.A, J. Engelberg, and W. Mullins, 2023, Echo Chambers, Review of Financial Studies 36(2): 450-500.

Cookson, J.A., R. Lu, W. Mullins, and M. Niessner, 2023, The social signal, University of Colorado working paper.

Cookson, J.A., and M. Niessner, 2019, Why don't we agree? Evidence from a social network of investors, Journal of Finance 75(1): 173-228.

Eisenstein, J., B. O'Connor, N.A. Smith, and E.P. Xing, 2014, Diffusion of lexical change in social media, PLoS One 9(11): 1-13.

Fama, E.F., and J.D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, Journal of Political Economy 81(3): 607-636.

Fang, L., and J. Peress, 2009, Media coverage and the cross-section of stock returns, Journal of Finance 64(5): 1985-2428.

Garcia, D., X. Hu, and M. Rohrer, 2023, The colour of finance words, Journal of Financial Economics 147(3): 525-549.

Giannini, R., P. Irvine, and T. Shu, 2019, The convergence and divergence of investors' opinions around earnings news: Evidence from a social network, Journal of Financial Markets, 42:94–120.

Gorodnichenko, Y., T. Pham, and O. Talavera, 2021, Central bank communication on social media: What, to whom, and how? University of Birmingham working paper.

Gu, C., and A. Kurov, 2020, Informational role of social media: Evidence from Twitter sentiment, Journal of Banking and Finance 121:105969.

Hu, D., C.M. Jones, V. Zhang, and X. Zhang, 2021, The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery, Peking University working paper.

Hutto, C., and E. Gilbert, 2014, Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* 8(1).

Irvine, P.J., S. Shen, and T. Shu, 2021, Aggregate attention, Texas Christian University working paper.

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

Kelley, E.K., and P.C. Tetlock, 2013, How wise are crowds? Insights from retail orders and stock returns, Journal of Finance 68(3): 1229-1265.

Loughran, T., and B. McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, Journal of Finance 66(1): 35-65.

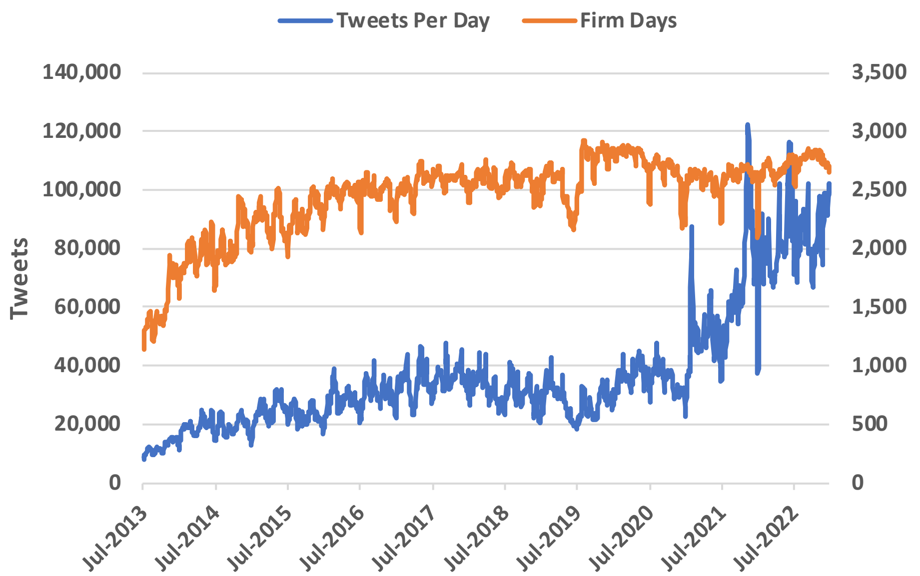
Mitchell, A., E. Shearer, and G. Stocking, 2021, News on Twitter: Consumed by most users and trusted by many. Pew Research Center Survey.

Newey, W. K., and K.D. West, 1987, Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 777-787.

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

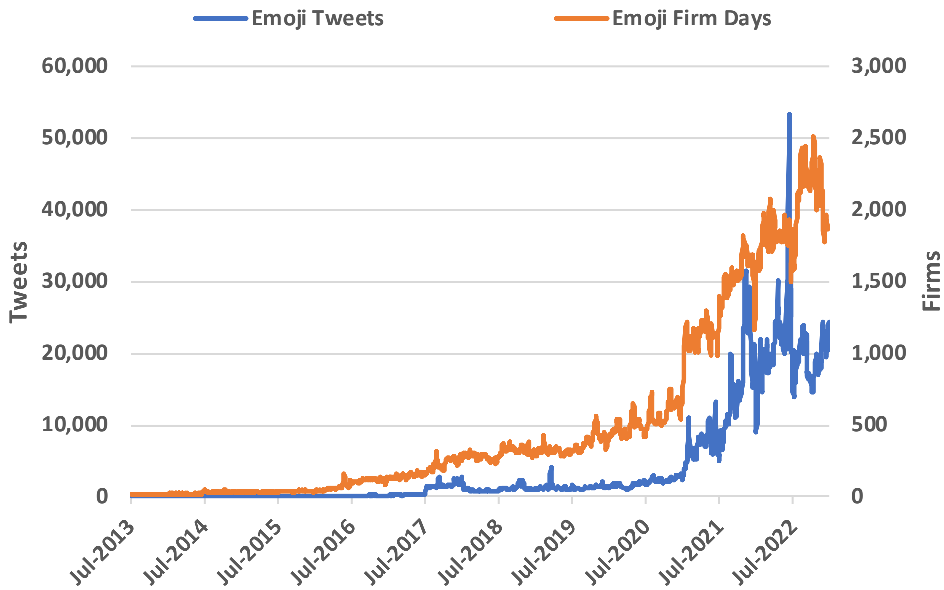
Tetlock, P.C., M. Saar-Tsechansky, and S. Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, Journal of Finance 63(3): 1437-1467.

Figure 1: Tweets and Firm Days for Entire Sample



This figure provides the number of tweets that reference a Russell 3000 firm with a cashtag ($) and the number of referenced firms each day. In blue we plot the total number of tweets each day, using a 5-day moving average. In orange we plot the number of firms per day with at least one tweet. We only include original tweets (i.e., no re-tweets) that are written in English and contain only one cashtag.

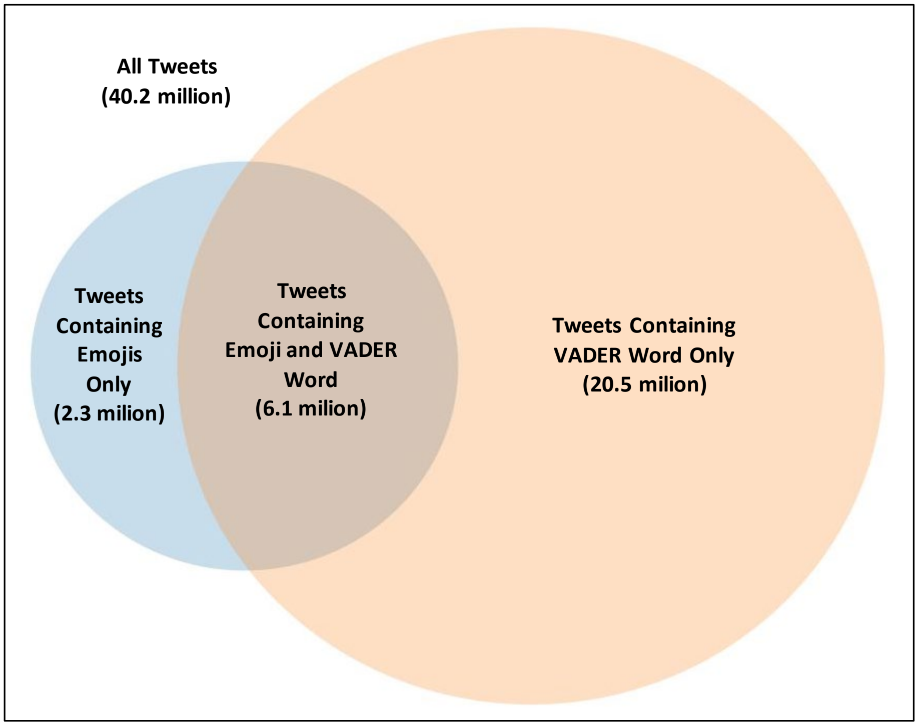
Figure 2: Emoji Tweets and Firm Days



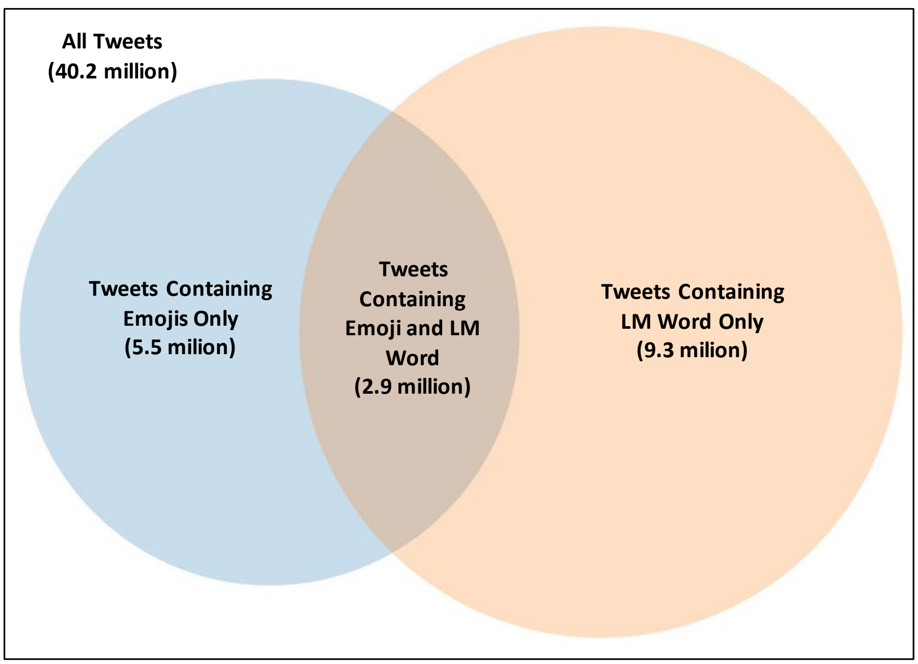
This figure provides the number of tweets containing an emoji that reference a Russell 3000 firm with a cashtag ($) and the number of referenced firms each day. In blue we plot the total number of tweets each day, that contain an emoji, using a 5-day moving average. In orange we plot the number of firms per day with at least one tweet that contained an emoji. We only include original tweets (i.e., no retweets) that are written in English and contain only one cashtag.

Figure 3: Scoring Methods Dictionary Coverage

Panel A: Emojis vs. VADER Words

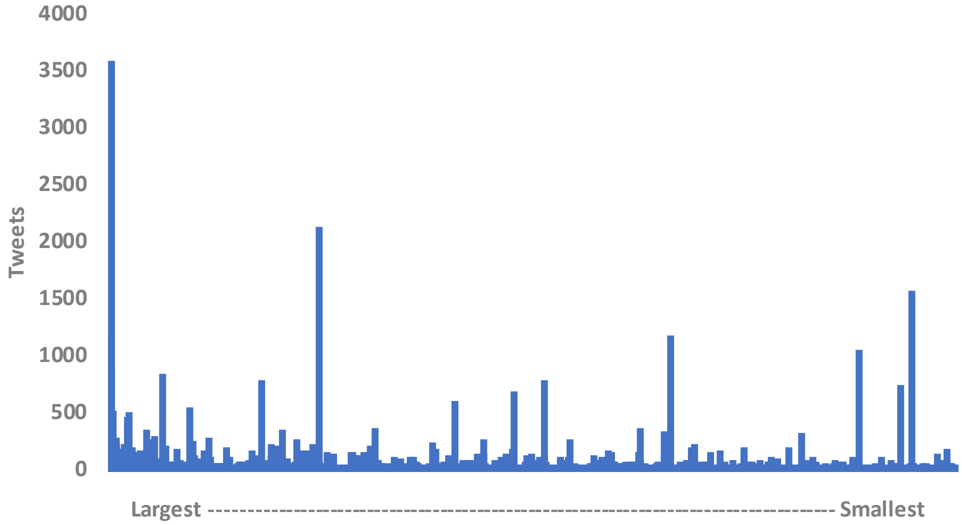


Panel B: Emojis vs. LM Words



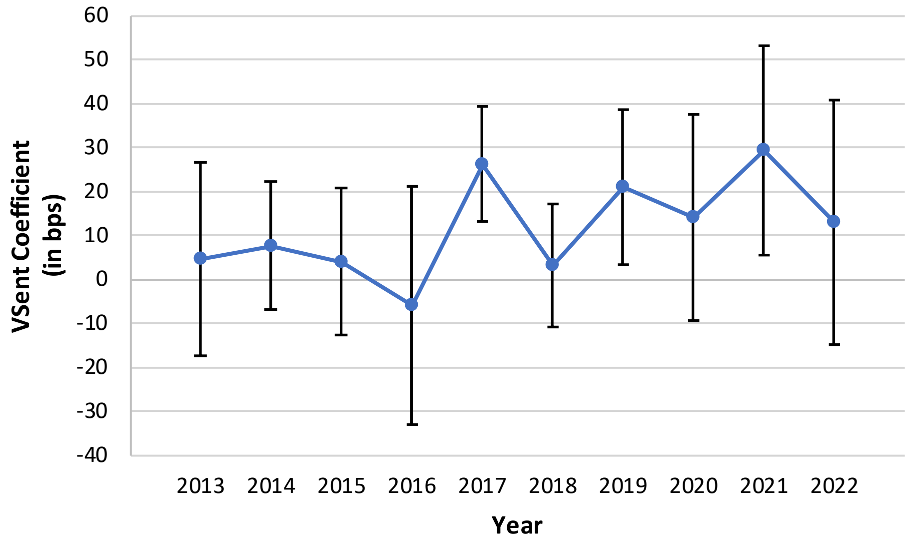
This figure presents Venn diagrams summarizing tweets that are scored by various dictionaries for the years 2021 and 2022. In Panel A we depict the coverage of the emoji dictionary and the VADER dictionary. In Panel B we depict the coverage of the emoji dictionary and the LM dictionary (any of the 2,692 positive/negative words). We also report the total number of tweets that contained a cashtag of a Russell 3000 firm in the upper left of each graph. We only include original tweets (i.e., no retweets) that are written in English and contain only one cashtag.

Figure 4: Distribution Of Daily Tweets By Firm Size



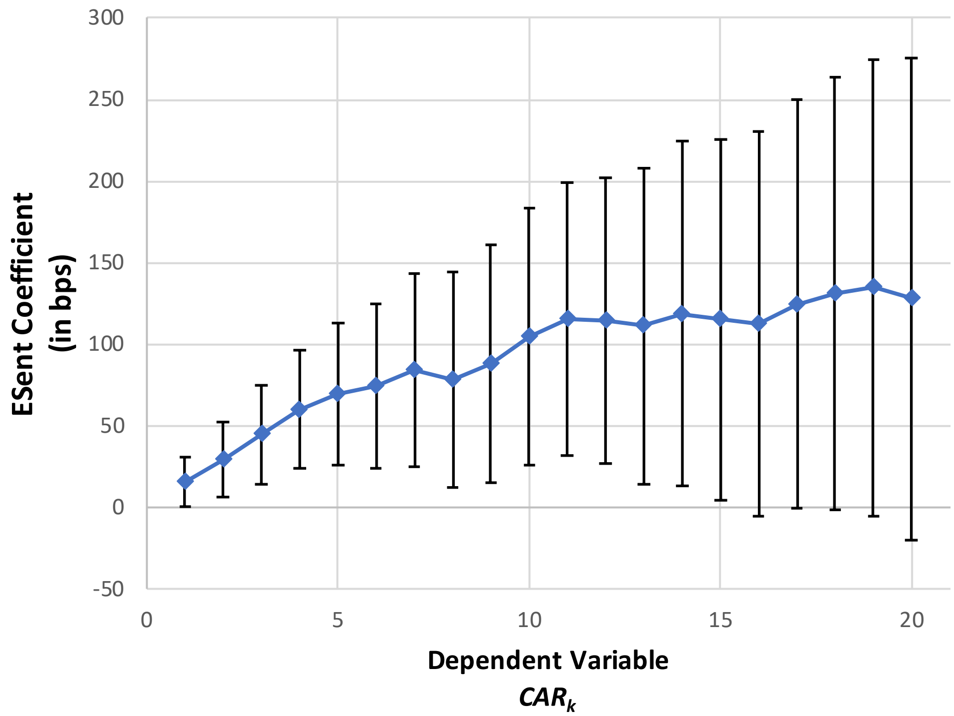
This figure provides the distribution of the number of daily tweets in 2022, based on firm size. We rank the firms according to market cap in December 2021 along the X-axis (largest to smallest from left to right) and plot each firm's average number of tweets per day over the year 2022. We only include original tweets (i.e., no retweets) that are written in English and contain only one cashtag.

Figure 5: *VSent* Coefficient with 95% CI



This figure provides the coefficient estimates from a pooled regression analysis cumulative abnormal returns (CAR) on the sentiment measure *VSent* over the years 2013 to 2022. We re-estimate the model from Column (4) in Table III Panel A each year separately. The figure contains yearly *VSent* coefficient estimates and their 95% confidence intervals.

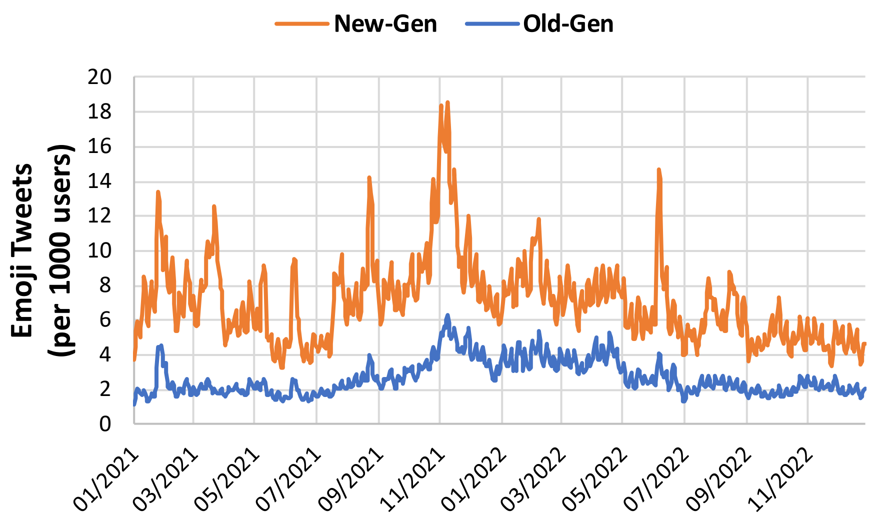
Figure 6: E*Sent* Coefficient with 95% CI



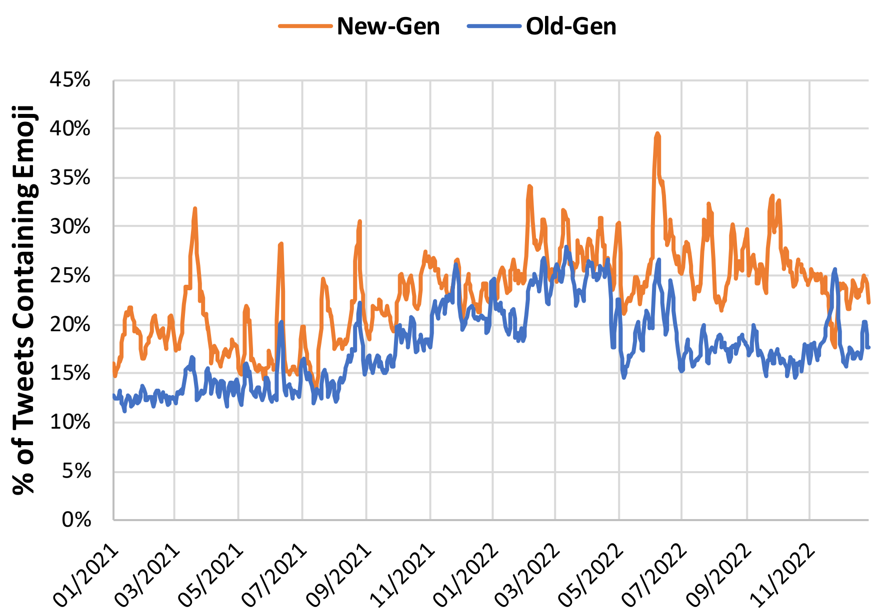
This figure provides the coefficient estimates from a pooled regression analysis of various cumulative abnormal returns (CAR) on the sentiment measure *ESent*. The model used for each regression is the same that is in column 1 of table 6. We estimate the model separately using $CAR_{t+1:t+k}$ for all k = 1 through 20 and report each *ESent* coefficient estimate along with its 95% confidence interval.

Figure 7: New-Gen vs. Old-Gen Emoji Usage

Panel A:

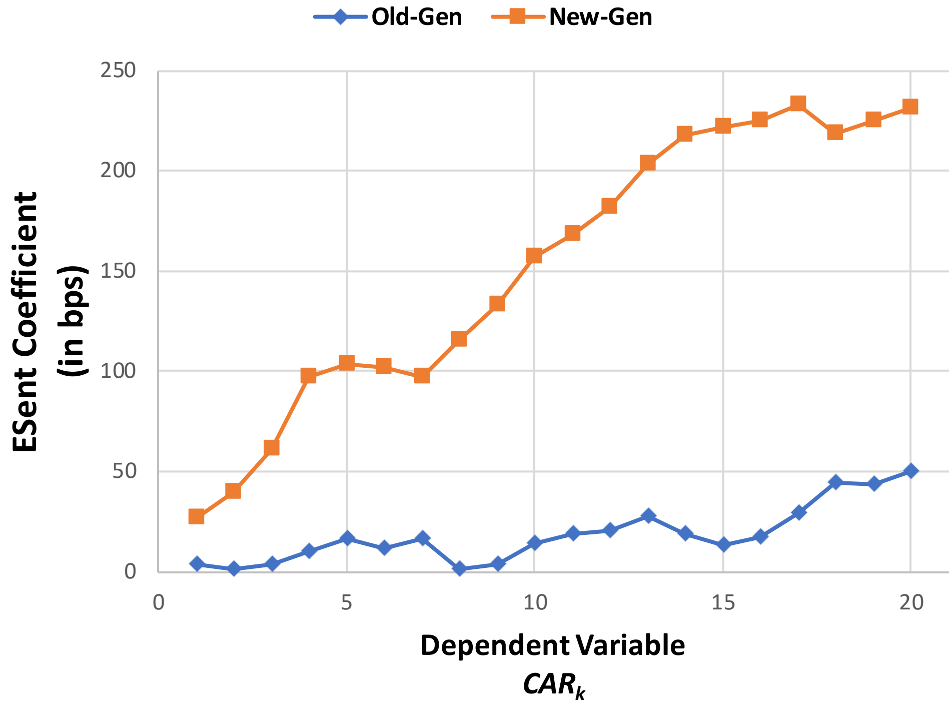


Panel B:



This figure depicts the emoji usage for two separate groups of authors during the years 2021 and 2022. We define "Old-Gen" authors as Twitter users who created their accounts prior to January 2020 and "New-Gen" authors as those who created their accounts during the year 2020. In Panel A we plot the number of daily emoji tweets per 1,000 users for both groups of authors. In Panel B we plot the percentage of daily tweets that contain an emoji for both the Old-Gen and the New-Gen authors. New-Gen users are depicted with orange lines, and Old-Gen users are depicted with blue lines.

Figure 8: New-Gen vs. Old-Gen *ESent* Coefficient



This figure provides the coefficient estimates from a pooled regression analysis of various cumulative abnormal returns (CAR) on the sentiment measure *ESent*. The models used for each regression are the same as those in Table 9, Columns (1) and (3). We compute the *ESent* variable separately for New-Gen and Old-Gen authors as defined above, and we estimate models separately using $CAR_{t+1:t+k}$ for all k = 1 through 20. We report the *ESent* coefficient estimates for each set of authors. New-Gen users are depicted with orange lines, and Old-Gen users are depicted with blue lines.

Table 1

Summary Statistics

This table provides summary statistics for variables observed at the firm-day level. We present average daily cross-sectional statistics within three different years: 2014, 2018, and 2022. Panel A contains tweet characteristic variables. The variable *NumTweets* is the number of tweets that contained a scorable word for a given firm on a given day. *NumWords* is the number of scorable words that were contained in a given tweet on a firm day. *NumEmojiTweets* is the number of tweets that contained a scorable Emoji for a given firm on a given day. *NumEmojis* is the number of scorable Emojis that were contained in a given tweet on a firm day. Panel B contains sentiment variables The variable *VSent* is the compound sentiment score that is output from the VADER model of Hutto and Gilbert (2014). *ESent* is the compound sentiment score derived from our new Emoji dictionaries. *LMNeg* is the percentage of negative words calculated based on Loughran and McDonald (2011). The statistics reported are the number of firms, mean, min, max, standard deviation, and the 5th, 25th, 50th, 75th, and 95th percentile.

	Panel A: Tweet Level										
	Year	N	Mean	Min	Max	STD	P5	P25	P50	P75	P95
	2014	1715	8.97	1.00	1032.52	37.04	1.00	1.30	2.82	6.50	28.40
NumTweets	2018	2583	12.85	1.00	1316.61	40.18	1.15	3.20	6.54	12.55	35.34
	2022	2646	27.62	1.00	5897.89	170.40	1.05	2.76	6.20	15.14	81.57
	2014	1715	8.82	1.00	27.74	4.50	2.55	5.70	8.05	11.30	17.65
NumWords	2018	2583	9.43	1.00	46.08	5.75	3.11	5.96	8.09	11.22	21.14
	2022	2646	13.81	1.00	54.96	9.45	2.69	6.83	11.42	18.54	33.46
	2014	18	2.48	1.00	12.23	2.88	1.00	1.00	1.25	2.61	7.02
NumEmojiTweets	2018	250	5.19	1.00	517.97	35.58	1.00	1.00	1.02	2.07	10.21
	2022	1660	10.84	1.00	3389.69	98.88	1.00	1.03	2.13	4.65	25.83
	2014	18	1.76	1.00	8.22	1.53	1.00	1.00	1.05	1.98	4.08
NumEmojis	2018	250	2.18	1.00	17.78	1.58	1.00	1.05	1.67	2.84	5.25
	2022	1660	2.47	1.00	56.56	2.17	1.00	1.05	1.72	3.09	6.37

	Panel B: Sentiment Variables										
	Year	N	Mean	Min	Max	STD	P5	P25	P50	P75	P95
	2014	503	13.21	3.78	98.10	8.68	5.17	7.90	11.01	15.81	28.26
LMNeg	2018	957	12.65	2.41	99.54	8.93	3.83	7.23	10.92	15.41	25.72
	2022	1553	9.11	1.88	100.00	7.12	2.71	4.67	7.47	11.60	20.29
	2014	1268	0.21	-0.95	0.96	0.41	-0.57	-0.12	0.32	0.48	0.75
VSent	2018	2195	0.24	-0.96	0.97	0.37	-0.48	0.08	0.30	0.48	0.75
	2022	2338	0.34	-0.98	0.99	0.42	-0.52	0.16	0.41	0.65	0.89
	2014	18	-0.11	-0.68	0.55	0.27	-0.50	-0.26	-0.14	0.03	0.32
ESent	2018	250	-0.04	-0.79	0.88	0.27	-0.46	-0.20	-0.07	0.13	0.44
	2022	1660	0.13	-0.99	0.98	0.24	-0.24	0.02	0.12	0.26	0.55

Table 2

Correlations for Twitter Sentiment

This table provides correlations for the key variables in the analysis each year from 2013 to 2022. Numbers in the table represent average daily cross-sectional correlations. The variables *VSent* and *LMNeg* are as defined in Table 1. *LMPos* is the percentage of positive words calculated based on Loughran and McDonald (2011). *Tone* is calculated by subtracting *LMNeg* from *LMPos*. The variable ret_0 is a return variable for the same day the sentiment variable is measured. $ret_{-5:-1}$ is a return variable for returns for the five days preceding the day the sentiment is calculated. $ret_{-26:-6}$ is a return variable for returns for the for days from 26 days prior to the measured sentiment to 6 days prior to the sentiment measurement date.

	Year	ret_0	$ret_{-5:-1}$	$ret_{-26:-6}$	*VSent*	*LMPos*	*Tone*
Average Correlations with *LMNeg*	2013	-0.017	-0.007	-0.014	-0.242	-0.028	-0.706
	2014	-0.035	-0.015	-0.013	-0.308	-0.042	-0.687
	2015	-0.037	-0.019	-0.019	-0.330	-0.016	-0.727
	2016	-0.047	-0.028	-0.012	-0.301	-0.026	-0.727
	2017	-0.035	-0.023	-0.004	-0.238	-0.032	-0.708
	2018	-0.038	-0.032	-0.009	-0.256	-0.024	-0.710
	2019	-0.036	-0.020	-0.011	-0.279	-0.016	-0.679
	2020	-0.022	-0.014	-0.009	-0.255	0.041	-0.683
	2021	-0.018	-0.007	-0.005	-0.258	0.060	-0.603
	2022	-0.022	0.001	-0.001	-0.255	-0.013	-0.674
		ret_0	$ret_{-5:-1}$	$ret_{-26:-6}$	*LMNeg*	*LMPos*	*Tone*
Average Correlations with *VSent*	2013	0.069	0.049	0.030	-0.242	0.274	0.352
	2014	0.075	0.066	0.039	-0.308	0.286	0.398
	2015	0.066	0.073	0.045	-0.330	0.262	0.404
	2016	0.066	0.109	0.032	-0.301	0.279	0.395
	2017	0.058	0.066	0.029	-0.238	0.295	0.366
	2018	0.063	0.046	0.013	-0.256	0.263	0.356
	2019	0.066	0.057	0.026	-0.279	0.272	0.377
	2020	0.055	0.034	0.004	-0.255	0.263	0.371
	2021	0.045	0.057	0.028	-0.258	0.246	0.358
	2022	0.039	0.052	0.029	-0.255	0.331	0.406

Table 3

Pooled Regression of Future Returns on Twitter Sentiment

This table provides results from Equation (1), which is a pooled regression analysis of returns on various sentiment measures and controls. All variables are observed at the firm-day level. The sample used for these regressions is all Russell 3000 firm days that had measured sentiment from tweets from 2013 to 2022. The dependent variables are cumulative abnormal returns (CARs) for various time horizons. The dependent variable $CAR_{i,t+k1:t+k2}$ is the sum of daily CRSP value-weighted market-adjusted returns over the interval ($t+k1$, $t+k2$), expressed in basis points. In Panel A, the dependent variables are CAR_{+1} and $CAR_{+2:+5}$. In Panel B, the dependent variables are $CAR_{+6:+10}$ and $CAR_{+11:+20}$. The variables *VSent* and *LMNeg* are as defined in Table 1. All control variables are defined Section II. All models include time fixed effect and Newey-West (1987) standard errors. The t-statistics are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Panel A:

Dependent variable	CAR_{+1}			$CAR_{+2:+5}$		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
VSent	5.076***		3.853**	15.715***		14.963***
	(3.215)		(2.382)	(4.007)		(3.752)
LMNeg		-0.514***	-0.409**		-0.659*	-0.2510
		(-3.280)	(-2.545)		(-1.867)	(-0.701)
$ret_{-5:-1}$	-65.508***	-65.022***	-65.564***	-118.146**	-116.066**	-118.179**
	(-3.000)	(-2.985)	(-3.003)	(-2.154)	(-2.116)	(-2.155)
$ret_{-26:-6}$	-13.422*	-13.346*	-13.473*	-53.135**	-52.672**	-53.167**
	(-1.814)	(-1.804)	(-1.822)	(-2.557)	(-2.531)	(-2.559)
Ln_ME	0.1010	0.1040	0.1170	1.3500	1.3080	1.3600
	(0.168)	(0.173)	(0.195)	(0.645)	(0.626)	(0.649)
Ln_BM	2.7660	2.7770	2.7660	5.9830	6.0280	5.9840
	(1.024)	(1.027)	(1.024)	(0.662)	(0.666)	(0.662)
ret_0	6.0610	6.5730	5.4220	-249.655***	-245.555***	-250.051***
	(0.135)	(0.147)	(0.121)	(-2.899)	(-2.856)	(-2.905)
Constant	-2.2800	-0.2160	-1.5630	-23.4690	-17.7910	-23.0280
	(-0.236)	(-0.023)	(-0.162)	(-0.697)	(-0.538)	(-0.685)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,755,962	1,755,962	1,755,962	1,748,776	1,748,776	1,748,776

Panel B:

Dependent variable	$CAR_{+6:+10}$			$CAR_{+11:+20}$		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
VSent	5.1880		6.9340	12.4470		13.1160
	(1.030)		(1.458)	(1.523)		(1.587)
LMNeg		0.3930	0.5830		-0.1370	0.2230
		(0.831)	(1.307)		(-0.213)	(0.348)
$ret_{-5:-1}$	-24.8360	-23.7730	-24.7570	-199.265***	-197.356***	-199.235***
	(-0.521)	(-0.498)	(-0.519)	(-3.008)	(-2.972)	(-3.008)
$ret_{-26:-6}$	-16.5660	-16.2620	-16.4920	52.8770	53.3520	52.9050
	(-0.633)	(-0.621)	(-0.631)	(1.064)	(1.074)	(1.065)
Ln_ME	1.9520	1.9050	1.9290	3.9440	3.8890	3.9350
	(0.761)	(0.744)	(0.751)	(0.799)	(0.790)	(0.797)
Ln_BM	5.1800	5.2000	5.1790	13.0720	13.1090	13.0720
	(0.470)	(0.471)	(0.469)	(0.626)	(0.627)	(0.626)
ret_0	-50.3330	-47.2990	-49.4100	-171.511**	-167.116**	-171.153**
	(-0.797)	(-0.748)	(-0.782)	(-2.384)	(-2.307)	(-2.380)
Constant	-31.2880	-29.8870	-32.3170	-64.6480	-60.4350	-65.0430
	(-0.760)	(-0.742)	(-0.787)	(-0.817)	(-0.778)	(-0.823)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,739,815	1,739,815	1,739,815	1,721,969	1,721,969	1,721,969

Table 4

Point in Time (PIT) Emoji Dictionaries

This table provides statistics of the number of emojis, and emoji blocks that were scored for each year from 2012 to 2022. Emoji blocks are defined as multiple emojis of at least two types that appear in a single uninterrupted string. For each year, we include the total number of emojis, and emoji blocks scored with at least 100 tweets. We also report the total number of tweets used to compute *ESent* for all Emojis and the average sentiment score assigned.

	Emojis			**Emoji Blocks**		
Year	Number Scored	Frequency	Average Sentiment	Number Scored	Frequency	Average Sentiment
2013	27	69,234	0.588	<10	-	-
2014	61	34,738	0.413	<10	-	-
2015	69	36,078	0.707	<10	-	-
2016	80	31,506	0.527	<10	-	-
2017	120	367,978	0.707	<10	-	-
2018	251	825,116	0.736	18	11,510	0.855
2019	337	851,264	0.728	44	38,130	0.692
2020	396	679,447	0.752	40	9,868	0.809
2021	675	2,562,610	0.946	182	53,244	1.137
2022	886	10,914,696	0.986	595	298,108	1.305

Table 5

Correlations for Emoji Sentiment

This table provides correlations between *ESent* and other variables of interest each year from 2013 to 2022. Numbers in the table represent average daily cross-sectional correlations. *ESent* is the compound sentiment score that is output from the VADER model's use of our new Emoji only dictionary. All other variables are as defined above.

	Year	ret_0	$ret_{-5:-1}$	$ret_{-26:-6}$	*VSent*	*LMNeg*	*LMPos*	*Tone*
Average Correlations with *ESent*	2013	0.010	0.022	0.049	0.151	-0.009	0.138	0.083
	2014	0.005	0.012	-0.004	0.124	-0.024	0.011	0.018
	2015	0.029	0.030	0.043	0.086	0.052	0.072	0.011
	2016	0.024	0.002	-0.029	0.145	-0.004	0.087	0.059
	2017	0.037	0.010	0.004	0.260	-0.116	0.160	0.188
	2018	0.056	0.014	-0.017	0.168	-0.149	0.128	0.195
	2019	0.054	0.033	-0.006	0.191	-0.181	0.122	0.207
	2020	0.047	0.027	0.006	0.246	-0.152	0.114	0.186
	2021	0.042	0.035	0.011	0.225	-0.123	0.101	0.153
	2022	0.035	0.028	0.020	0.149	-0.102	0.066	0.110

Table 6

Pooled Regression of Future Returns on Emoji Sentiment

This table provides results from Equation (1) which is a pooled regression analysis of returns on various sentiment measures. All variables are observed at the firm-day level. The sample used for these regressions is all Russell 3000 firm days that had measured sentiment from tweets from 2021 to 2022. The dependent variables are cumulative abnormal returns (CARs) for various time horizons. The dependent variable $CAR_{i,t+k1:t+k2}$ is the sum of daily CRSP value-weighted market-adjusted returns over the interval ($t+k1$, $t+k2$), expressed in basis points. In Panel A, the dependent variables are CAR_{+1} and $CAR_{+2:+5}$. In Panel B, the dependent variables are $CAR_{+6:+10}$ and $CAR_{+11:+20}$. All return variables are in basis points. The variable *ESent* is the compound sentiment score that is output from the VADER model's use of our new Emoji only dictionary. All other variables are as defined above. All models include time fixed effect and Newey-West (1987) standard errors. The t-statistics are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Panel A

Dependent variable	CAR_{+1}			$CAR_{+2:+5}$		
	(1)	(2)	(3)	(4)	(5)	(6)
ESent	15.724**	14.544*	12.947*	52.599***	46.142**	48.710**
	(2.019)	(1.851)	(1.676)	(2.688)	(2.352)	(2.454)
VSent		7.1620			37.070**	
		(0.993)			(2.154)	
LMNeg			-2.487***			-3.402*
			(-3.790)			(-1.918)
$Ret_{-5:-1}$	-30.4590	-32.4730	-30.1880	-116.9520	-120.7100	-116.5950
	(-0.722)	(-0.762)	(-0.715)	(-1.019)	(-1.046)	(-1.016)
$Ret_{-26:-6}$	-1.7150	-1.2900	-1.7800	-43.7090	-42.5170	-43.7690
	(-0.124)	(-0.093)	(-0.129)	(-1.092)	(-1.054)	(-1.095)
Ln_ME	1.1960	1.3840	1.3900	6.9190	7.2890	7.1810
	(0.763)	(0.860)	(0.875)	(1.403)	(1.423)	(1.431)
Ln_BM	12.322*	12.131*	12.128*	35.245*	36.289*	34.962*
	(1.868)	(1.816)	(1.844)	(1.753)	(1.785)	(1.746)
Ret_0	-56.4200	-58.7400	-57.8730	-158.7840	-161.3350	-160.5950
	(-0.586)	(-0.607)	(-0.601)	(-1.390)	(-1.414)	(-1.408)
Constant	-24.2350	-29.4900	-23.5570	-129.8170	-147.104*	-128.7890
	(-0.933)	(-1.057)	(-0.909)	(-1.594)	(-1.683)	(-1.587)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	232,261	226,532	232,233	228,230	222,623	228,202

Panel B

Dependent variable	$CAR_{+6:+10}$			$CAR_{+11:+20}$		
	(1)	(2)	(3)	(4)	(5)	(6)
ESent	32.0410	15.3730	30.8310	20.8080	3.9140	15.0940
	(1.360)	(0.637)	(1.299)	(0.453)	(0.097)	(0.337)
VSent		43.732**			85.631*	
		(1.963)			(1.926)	
LMNeg			-1.2260			-5.2480
			(-0.719)			(-1.621)
$Ret_{-5:-1}$	55.4400	54.5180	55.5110	-42.0270	-41.3240	-41.6350
	(0.571)	(0.565)	(0.572)	(-0.327)	(-0.324)	(-0.324)
$Ret_{-26:-6}$	-20.6780	-19.4020	-20.7450	89.8390	92.3850	90.2080
	(-0.427)	(-0.400)	(-0.428)	(0.674)	(0.683)	(0.677)
Ln_ME	11.048*	11.419*	11.146*	24.268**	25.630**	24.709**
	(1.864)	(1.875)	(1.855)	(2.384)	(2.418)	(2.393)
Ln_BM	56.431**	57.060**	56.356**	108.162**	108.466**	107.752**
	(2.276)	(2.270)	(2.279)	(2.292)	(2.302)	(2.291)
Ret_0	116.2140	112.5480	115.6720	-41.3710	-61.0090	-45.6330
	(1.082)	(1.048)	(1.078)	(-0.251)	(-0.379)	(-0.277)
Constant	-210.119**	-228.635**	-209.846**	-448.406***	-496.542***	-447.480***
	(-2.158)	(-2.190)	(-2.162)	(-2.658)	(-2.686)	(-2.665)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	223,496	217,957	223,468	214,500	209,046	214,473

Table 7

Pooled Regression of Future Returns on Emoji Sentiment (with event controls)

This table provides results from pooled regressions similar to those in Table 6, but augmented to include additional information variables. The variables $DJNWNum\_Stories_{it}$ and $DJNWSent_{it}$ represent the number of individual stories and their average sentiment for the day respectively, and $DJNWSent_{it}$ equals zero if no stories appear on day t. The variables $EADate_{it}$ and $8\text{-}KDate_{it}$ are indicators set to one if day t contains an earnings release or 8-K, respectively, and zero otherwise. All other variables are as defined above. All models include time fixed effect and Newey-West (1987) standard errors. The t-statistics are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Panel A

Dependent variable	***CAR*** $_{+1}$			***CAR*** $_{+2:+5}$		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
ESent	15.323*	13.590*	12.818*	52.853***	46.369**	49.124**
	(1.959)	(1.720)	(1.653)	(2.729)	(2.407)	(2.499)
VSent		6.7600			33.428**	
		(0.981)			(2.091)	
LMNeg			-2.243***			-3.323*
			(-3.461)			(-1.892)
EAdate	19.8410	19.0180	19.3190	13.4210	15.7340	12.6530
	(1.608)	(1.558)	(1.566)	(0.625)	(0.734)	(0.590)
*EAdate**Ret_0*	-99.8580	-103.2240	-102.6180	-160.1630	-154.7830	-164.2250
	(-0.523)	(-0.538)	(-0.539)	(-0.604)	(-0.584)	(-0.619)
*8-K Date**Ret_0*	286.890**	284.643**	285.878**	293.383*	288.377*	291.915*
	(2.194)	(2.180)	(2.191)	(1.686)	(1.653)	(1.677)
8-K Date	-17.935**	-17.852**	-18.009**	23.284*	23.091*	23.175*
	(-2.277)	(-2.258)	(-2.286)	(1.834)	(1.816)	(1.830)
DJNW Num_Stories	0.0000	0.0120	0.0130	-0.2900	-0.2800	-0.2710
	(-0.001)	(0.068)	(0.072)	(-1.041)	(-1.008)	(-0.977)
DJNW Sent	40.598***	40.165***	39.777***	5.4850	5.2120	4.2660
	(8.207)	(8.195)	(8.109)	(0.613)	(0.581)	(0.483)
Ret $_{-5:-1}$	-31.8860	-31.4890	-31.6190	-116.7790	-117.5920	-116.3720
	(-0.754)	(-0.742)	(-0.747)	(-1.020)	(-1.021)	(-1.016)
Ret $_{-26:-6}$	-2.0120	-1.6310	-2.0160	-43.7020	-43.3160	-43.6920
	(-0.146)	(-0.118)	(-0.146)	(-1.093)	(-1.082)	(-1.094)
Ln_ME	0.9980	1.0950	1.1680	7.1180	7.3140	7.3730
	(0.630)	(0.677)	(0.729)	(1.444)	(1.455)	(1.471)
Ln_BM	12.143*	12.053*	11.956*	35.116*	35.219*	34.836*
	(1.845)	(1.838)	(1.821)	(1.755)	(1.764)	(1.748)
Ret $_0$	-86.6980	-87.7300	-87.5020	-171.6870	-179.1540	-172.9100
	(-0.840)	(-0.848)	(-0.848)	(-1.461)	(-1.525)	(-1.471)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	232,261	230,350	232,260	228,230	226,361	228,229

Table 7 (Continued)

Panel B

Dependent variable	*CAR +6:+10*			*CAR +11:+20*		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
ESent	30.8770	17.3500	29.6010	22.3880	3.8000	16.8300
	(1.317)	(0.719)	(1.254)	(0.494)	(0.094)	(0.381)
VSent		38.257*			77.873*	
		(1.789)			(1.896)	
LMNeg			-1.1410			-5.0080
			(-0.676)			(-1.566)
EAdate	-13.4710	-12.8410	-13.7270	75.327***	76.118***	74.181***
	(-0.629)	(-0.596)	(-0.643)	(3.097)	(3.129)	(3.058)
*EAdate**Ret_0*	-61.3280	-67.6420	-62.6480	-24.2590	-25.2930	-29.8990
	(-0.261)	(-0.288)	(-0.268)	(-0.116)	(-0.120)	(-0.143)
*8-K Date**Ret_0*	-41.7300	-41.1220	-42.5820	71.8030	61.1430	67.8120
	(-0.151)	(-0.148)	(-0.154)	(0.353)	(0.302)	(0.333)
8-K Date	34.512***	35.471***	34.470***	31.1220	31.310*	30.9270
	(3.114)	(3.176)	(3.118)	(1.644)	(1.661)	(1.640)
DJNW Num_Stories	-0.489*	-0.459*	-0.483*	-0.5580	-0.5170	-0.5290
	(-1.859)	(-1.729)	(-1.842)	(-1.363)	(-1.297)	(-1.287)
DJNW Sent	5.4590	5.6520	5.0400	26.387*	27.185*	24.5470
	(0.520)	(0.543)	(0.485)	(1.708)	(1.766)	(1.636)
Ret-5:-1	55.8190	54.5270	55.9550	-41.7440	-46.7510	-41.1240
	(0.575)	(0.564)	(0.576)	(-0.325)	(-0.366)	(-0.320)
Ret-26:-6	-20.6600	-20.8310	-20.6710	89.7040	90.8950	89.8600
	(-0.427)	(-0.431)	(-0.427)	(0.672)	(0.677)	(0.673)
Ln_ME	11.369*	11.605*	11.458*	24.580**	25.477**	24.960**
	(1.913)	(1.905)	(1.903)	(2.399)	(2.412)	(2.403)
Ln_BM	56.376**	56.102**	56.287**	107.988**	108.202**	107.531**
	(2.277)	(2.271)	(2.279)	(2.293)	(2.321)	(2.292)
Ret0	121.0260	115.8680	120.6580	-53.9260	-73.0260	-55.5930
	(1.044)	(0.998)	(1.041)	(-0.300)	(-0.416)	(-0.310)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes
Observations	223,496	221,643	223,495	214,500	212,660	214,499

Table 8

Future Returns on Emoji Sentiment by Author Generation

This table provides results from pooled regressions similar to those in Table 6, except we recompute the *ESent* and *VSent* variables separately for two cohorts of Twitter users. We define "Old-Gen" authors as Twitter users who created their accounts prior to January 2020 and "New-Gen" authors as those who created their accounts during the year 2020. The sample used for these regressions is all Russell 3000 firm days that had measured sentiment from tweets from 2021 to 2022. All variables are as defined above. All models include time fixed effect and Newey-West (1987) standard errors. The t-statistics are reported in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01

Panel A

Dependent variable	***CAR*** $_{+1}$				***CAR*** $_{+2:+5}$			
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Esent (New)	27.119*	23.7220			72.392**	72.335**		
	(1.767)	(1.509)			(2.126)	(2.139)		
Esent (Old)			3.9810	0.9900			12.0100	-0.0670
			(0.462)	(0.120)			(0.576)	(-0.004)
Vsent(New)		7.4210				1.3950		
		(0.901)				(0.072)		
Vsent(Old)				10.6660				38.853**
				(1.626)				(2.358)
Ret $_{-5:-1}$	15.9530	16.2560	-2.3180	-4.6010	-215.3800	-211.2000	-134.9710	-137.8710
	(0.195)	(0.197)	(-0.040)	(-0.078)	(-1.078)	(-1.049)	(-0.878)	(-0.895)
Ret $_{-26:-6}$	-13.5530	-14.9890	-8.4510	-8.6790	-78.383**	-77.869**	-78.334*	-80.097*
	(-0.817)	(-0.909)	(-0.567)	(-0.583)	(-1.989)	(-1.962)	(-1.880)	(-1.933)
Ln_ME	2.3380	2.5100	2.0180	2.1290	5.3060	5.3450	5.0410	5.3900
	(1.362)	(1.423)	(1.334)	(1.387)	(0.906)	(0.897)	(1.002)	(1.060)
Ln_BM	22.081**	22.820**	21.055***	20.913***	66.628*	68.692*	50.465*	50.260*
	(2.015)	(2.034)	(2.638)	(2.615)	(1.805)	(1.860)	(1.950)	(1.937)
Ret $_{0}$	-143.8160	-148.8610	-70.9900	-74.1890	-78.6940	-84.4430	-163.0390	-167.8010
	(-0.970)	(-1.007)	(-0.633)	(-0.659)	(-0.621)	(-0.667)	(-1.556)	(-1.606)
Constant	-53.350*	-58.284*	-41.9720	-46.681*	-129.3120	-132.4810	-107.9300	-124.1490
	(-1.802)	(-1.862)	(-1.629)	(-1.745)	(-1.272)	(-1.258)	(-1.266)	(-1.417)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	60,872	59,522	113,363	112,531	60,565	59,220	112,617	111,790

Panel B

Dependent variable	***CAR*** $_{+6:+10}$				***CAR*** $_{+11:+20}$			
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Esent (New)	57.8600	38.6760			74.2030	78.6650		
	(1.438)	(1.039)			(0.819)	(0.987)		
Esent (Old)			0.0890	-5.4340			38.1940	20.9210
			(0.003)	(-0.229)			(0.715)	(0.422)
Vsent(New)		29.5880				34.8550		
		(1.030)				(0.760)		
Vsent(Old)				16.2410				50.5980
				(0.722)				(1.372)
Ret $_{-5:-1}$	-55.5990	-54.9130	-29.4430	-28.4360	67.5870	74.3330	-20.4920	-22.5090
	(-0.615)	(-0.608)	(-0.320)	(-0.310)	(0.515)	(0.557)	(-0.153)	(-0.168)
Ret $_{-26:-6}$	-10.4970	-13.0880	-30.7290	-30.0850	278.5140	283.2210	151.9320	152.8010
	(-0.173)	(-0.214)	(-0.576)	(-0.564)	(1.225)	(1.236)	(0.851)	(0.852)
Ln_ME	8.3540	8.6930	7.9730	8.2850	18.8890	19.9070	18.522*	19.154*
	(1.242)	(1.274)	(1.312)	(1.340)	(1.574)	(1.605)	(1.683)	(1.715)
Ln_BM	83.326*	84.304**	68.585**	68.102**	211.917***	209.368***	137.491**	136.555**
	(1.935)	(1.969)	(2.205)	(2.184)	(2.691)	(2.678)	(2.342)	(2.324)
Ret $_{0}$	140.4650	141.5420	91.5280	87.7340	-120.6990	-124.5660	-114.8460	-121.8520
	(1.112)	(1.109)	(0.937)	(0.890)	(-0.675)	(-0.707)	(-0.739)	(-0.797)
Constant	-194.009*	-209.357*	-167.3850	-177.096*	-431.375**	-462.210**	-379.985**	-404.496**
	(-1.676)	(-1.711)	(-1.633)	(-1.649)	(-2.081)	(-2.065)	(-2.034)	(-2.087)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newey-West SE's	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	60,145	58,810	111,619	110,796	59,269	57,951	109,507	108,694