

Social Media Emotions and Market Behavior*

Domonkos F. Vamossy[†]

April 8, 2024

Abstract

I explore the relationship between investor emotions expressed on social media and asset prices. The field has seen a proliferation of models aimed at extracting firm-level sentiment from social media data, though the behavior of these models often remains uncertain. Against this backdrop, my study employs EmTract, an open-source emotion model, to test whether the emotional responses identified on social media platforms align with expectations derived from controlled laboratory settings. This step is crucial in validating the reliability of digital platforms in reflecting genuine investor sentiment. My findings reveal that firm-specific investor emotions behave similarly to lab experiments and can forecast daily asset price movements. These impacts are larger when liquidity is lower or short interest is higher. My findings on the persistent influence of sadness on subsequent returns, along with the insignificance of the one-dimensional valence metric, underscores the importance of dissecting emotional states. This approach allows for a deeper and more accurate understanding of the intricate ways in which investor sentiments drive market movements.

Keywords: Deep Learning; Investor Emotions; Text Analysis; Social Media; Return Predictability.

JEL Codes: G41; L82.

*I am extremely grateful to Stefania Albanesi, Lee Dokyun, Sera Linardi, and Zhi Da for their continued guidance and support throughout this project. Special thanks to StockTwits for sharing their data. This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided.

[†]Department of Economics, University of Pittsburgh, d.vamossy@pitt.edu.

1 Introduction

The intricate relationship between human emotions and markets is a foundational aspect of economic theory, highlighted by John Maynard Keynes’s concept of “animal spirits”. He argued that it is the spontaneous optimism, not grounded in mathematical expectations, that often drives decisions:

Even apart from the instability due to speculation, there is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than on a mathematical expectation, whether moral or hedonistic or economic. Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as a result of animal spirits – of a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities. Keynes [1936]

This quote captures the essence of the human element in economic dynamics, where sentiment and instinct can outweigh rational calculations. Akerlof and Shiller [2010] further expanded on this in, arguing that understanding the emotional and psychological drivers is crucial for navigating financial markets effectively. They suggest that the economy’s functioning is significantly influenced by confidence, fairness, corruption, and money illusion, all facets of Keynes’s original “animal spirits.”

In this paper, I examine the link between firm-specific investor emotions and the movements in their asset prices. My research is guided by two main questions: (1) Can firm-specific investor emotions predict daily price changes? and (2) Are the emotions derived from social media consistent with those observed in laboratory experiments? The challenge of accurately measuring investor emotions has led researchers to rely on indirect proxies or limit their analysis to experimental settings. However, by leveraging a large and unique dataset alongside modern advancements in text analysis, I have managed to navigate the

complexities associated with studying investor emotions. Specifically, I utilize data from StockTwits, a social media platform where investors exchange views on stocks, which provides firm-specific posts, enabling the calculation of emotions. My analysis incorporates an extensive dataset comprising over 88 million messages collected from 2010 to September 2021.

This paper leverages Vamossy and Skog [2023], which employs deep learning techniques and an extensive dataset of investor communications. Specifically, the Vamossy and Skog [2023] model processes text from social media and assigns to each message a set of variables that represent seven distinct emotional states: neutral, happy, sad, anger, disgust, surprise, and fear. This approach follows the emotional framework outlined by Breaban and Noussair [2018]. The emotional variables I work with are probabilistic, meaning that the sum of the seven emotions for any given message equals 1. For example, the text “not feeling it :)” would yield an emotion tuple like (0.064, 0.305, 0.431, 0.048, 0.03, 0.038, 0.084), corresponding to the aforementioned emotional states. Changing the emoticon from :) to :(in “not feeling it :(” alters the model’s output significantly, demonstrating how the model adeptly integrates emoticons to refine the emotional analysis of the text.¹ Using this tool, I can investigate whether investor emotions specific to a firm, captured before the market opens, can predict the firm’s daily price movements. My approach involves applying the tool to quantify the emotional content of each message through textual analysis, and then averaging these results for each firm by day, distinguishing between messages at market open and close. Beyond measuring emotional content, I also categorize messages using two classification schemes: one that differentiates messages related to earnings, firm fundamentals, or stock trading from general discussions, and another that separates original content from reposted information.

To analyze whether these emotions can predict daily price movements, I employ a fixed effects model that leverages within-firm emotional variations. Using within-firm variation shuts down some mechanisms, such as the possibility that consistently high-return firms

¹For additional examples, visit the interactive interface.

might elicit consistently positive investor emotions. By including firm-specific fixed effects, I can ensure that my findings aren't merely reflections of this. Additionally, I control for calendar day fixed effects to account for any broad market influences that might simultaneously affect all firms.

To address potential estimation concerns, I take several steps. For example, I focus on emotions expressed before the market opens — between 4:00 PM the previous day and 9:29 AM on the trading day — to establish a clear temporal separation between the emotions and the subsequent trading behavior. This helps in excluding reactionary emotions. To tackle the risk of misattribution, where the emotions captured might not accurately reflect the true sentiment, Vamossy and Skog [2023] validate the emotion measures by comparing them with self-tagged bullish or bearish texts. I add to this by examining the link between emotion measures and contemporaneous asset prices and conducting robustness checks with an alternative emotion model. My findings suggest that assets that have increased in value tend to be associated with messages classified as happier, and these observations remain consistent even when using the alternative emotion model.

My analysis focuses on understanding how investor emotions influence daily asset price movements, leading to three key findings. First, I document that firm-specific investor emotions can predict the firm's daily price movements. For example, I find that variations in investor enthusiasm are associated with slightly higher daily returns. This effect is evident in both messages that provide new information and those that circulate existing information. When I compare messages related to earnings, firm fundamentals, or stock trading with those containing other types of information, the latter seems to have a slightly larger impact on daily returns. To confirm the relationship between daily price movements and investor emotions, I also explore alternative emotion variables from emotion metadata compiled by other researchers. My findings remain significant, showing comparable point estimates. Moreover, I observe that the effects of emotions are more pronounced during periods of elevated short interest and lower liquidity. Second, emotions expressed on social media align with those

expected from controlled laboratory experiments. This consistency strengthens the case for using social media as a viable proxy for investor sentiment, indicating that the emotional expressions captured online are indicative of wider investor behavior patterns. Last, the persistence of sadness on subsequent returns and the insignificance of the combined valence metric underscores the importance of dissecting emotional states to gain a deeper and more accurate insight into how investor sentiments drive market movements.

My research contributes to the behavioral finance field by extending the understanding of the link between market behavior and emotional states. Unlike previous studies that often relied on indirect proxies, such as Kamstra et al. [2003] linking seasonal mood changes to market returns or Hirshleifer and Shumway [2003] correlating weather with stock performance, my approach directly measures investor emotions through social media posts related to specific stocks, addressing concerns about the relevance of emotional measures to the firms. This direct method of capturing investor sentiment allows for a clearer connection between emotions and stock prices, sidestepping potential confounders highlighted by critiques like Jacobsen and Marquering [2008]. While research in this area has predominantly focused on market-level emotions and index returns², with few exceptions like Li et al. [2016] and Long et al. [2021] exploring firm-specific sentiments, my work adds a novel dimension by showing that investor emotions at the firm level can predict daily stock price movements, thereby offering a more granular perspective on how sentiment influences market dynamics.

The connection between market behavior and emotional states has been extensively examined in controlled laboratory settings, where researchers have investigated how emotions contribute to the formation of bubbles in experimental asset markets. For example, Breaban and Noussair [2018] utilized facial expressions of traders to measure emotions, finding that positive emotions correlate with higher prices and bigger bubbles, whereas pre-market fear among traders tends to result in lower prices. Andrade et al. [2016] also observed that bub-

²Bollen et al. [2011] discovered that the overall mood on Twitter could forecast movements in the stock market, whereas Gilbert and Karahalios [2010] identified that the degree of anxiety in Live Journal blog posts was indicative of forthcoming declines in stock prices.

bles tend to be larger when induced investor enthusiasm is higher. My contribution to this field is demonstrating that emotions reflected in investor messages on social media platforms mirror those observed in laboratory experiments, suggesting a parallel in how emotions influence market dynamics both in controlled settings and in the real world via observational data. Nevertheless, it's crucial to recognize that the observational nature of social media data, while insightful, lacks the precision and control over variables that laboratory experiments offer. This inherent limitation complicates the task of drawing clear causal inferences from the data observed in such environments.

Another contribution of my paper is to the body of research examining social media's impact on capital markets. The relevance of social media in this domain is underscored by studies such as Blankespoor et al. [2014], who demonstrate that firms can diminish information asymmetry among investors by using Twitter to widely distribute news, press releases, and other disclosures. Jung et al. [2018] report that a significant portion of S&P 1500 firms have established a corporate presence on social media platforms like Facebook and Twitter. Additionally, there's a growing body of work, including research on StockTwits and Twitter, that investigates how investors' engagement with various online platforms, from search engines and financial websites to forums, influences market dynamics. This line of inquiry has yielded mixed results on the predictive power of online information for future earnings and stock returns. For instance, Da et al. [2011] use Google search volume as an indicator of investors' information demand, finding that increased searches foreshadow short-term stock price increases followed by a reversal within a year. Drake et al. [2012] observe that the correlation between returns and earnings weakens when Google searches spike before earnings announcements. Studies by Antweiler and Frank [2004] and Das and Chen [2007] link the volume of message board posts to stock return volatility but not to the returns themselves. Chen et al. [2014] show that user-generated content on the Seeking Alpha platform can forecast earnings and long-term stock returns post-report publication. The literature also delves into how social media activity around earnings announcements affects

investor behavior, with Curtis et al. [2014] finding a connection between social media buzz and heightened sensitivity to earnings news and surprises, and Cookson and Niessner [2020] indicating that while StockTwits discourse may not directly impact market movements, the platform’s message disagreements are a reliable predictor of unusual trading volumes.³

The remainder of the paper is organized as follows. Section 2 describes the data; Section 3 presents my primary results; Section 4 explores heterogeneous effects and conducts a sensitivity analysis. Section 5 concludes.

2 Data

2.1 Social Media Data (StockTwits)

My dataset on investor emotions is sourced from StockTwits, established in 2008 as a platform for investors to exchange opinions on stocks. The interface of StockTwits is reminiscent of Twitter, allowing users to post messages limited to 140 characters, which was expanded to 280 characters in late 2019. Users employ “cashtags” followed by a stock’s ticker symbol (for example, \$AMZN) to associate their messages with specific companies. While StockTwits does not offer direct integration with other social media networks, it does enable users to share their posts on personal accounts across Twitter, LinkedIn, and Facebook.

My initial dataset spans from 1 January 2010 to 30 September 2021, encompassing a total of 242 million messages. For each message, I can access sentiment indicators marked by users (labeled as bullish, bearish, or unclassified), a sentiment score calculated by StockTwits, “cashtags” linking messages to specific stocks, the number of likes a message receives, and a unique user identifier. This identifier enables me to delve into user characteristics, such as the number of followers they have. For the majority of users, there is also data on their self-declared investment philosophy, which can be categorized in two ways: (1) by their Approach - including technical, fundamental, momentum, value, growth, and global macro

³Additional studies by Vamossy [2023] and Vamossy [2021].

strategies; and (2) by their Holding Period - classified as a day trader, swing trader, position trader, or long-term investor. To analyze variations among investors, I combine technical with momentum and value with fundamental approaches and differentiate between short- and long-term investment horizons. Additionally, users disclose their level of experience, which ranges from novice, through intermediate, to professional. This detailed information on users' investment styles, experience levels, and strategies is instrumental in examining the diversity of investor emotions on the platform.

I eliminate messages that seem to be automated, defining these as any cleaned message posted more than 100 times by the same user throughout the period from 2010 to 2021. My focus is on messages explicitly linked to specific stocks, leading me to concentrate on those mentioning only one ticker symbol. Additionally, I set a threshold requiring a minimum of ten posts per stock for each trading session (including both market and non-market hours) to minimize the impact of random noise. I also exclude inactive tickers as defined by SECSTAT and limit my study to common ordinary shares listed on U.S. exchanges. These sample restrictions are detailed in Table 1. I end up with 88M messages authored by 984,434 users, covering 4,319 tickers.

In Panel (a) of Figure 1, I plot the average word count of messages over time, noting a consistent trend with a notable increase in late 2019 due to the expansion of the character limit from 140 to 280. The average message length reaches a high of 20 words, and 97.5% of the messages in my dataset contain 64 words or fewer. Given that I use only the first 64 words of each message to identify the underlying emotions, this cap is unlikely to compromise the accuracy of my analysis.

In Panel (b) of Figure 1, I present the trend of message volume over time in my dataset, showing a significant increase in the early stages, a leveling off around 2017, and then a marked rise in 2020. To account for the evolving nature of both the sample size and the content of posts, I incorporate time fixed effects in my analysis.

I further investigate the timing of investors' posts, specifically whether they align with

the influx of daily news, reflecting real-time adjustments in beliefs, or if they tend to occur in the evening, suggesting a more reflective, analytical approach after work hours. In Panels (c) and (d) of Figure 1, I display the distribution of messages by both the day of the week and the hour of the day. The bulk of activity on the platform coincides with market hours (Monday to Friday, 9 AM to 4 PM), indicating that investors are likely reacting and adjusting their perspectives in response to live market developments.

2.2 Short Interest and Pricing

I constructed additional variables using conventional data sources. Variables related to stock prices were sourced from the merged CRSP/COMPUSTAT file, while short interest data came from Compustat’s Supplemental Short Interest File. I then matched this information with the StockTwits data based on the ticker symbols and dates.

2.3 Descriptive Statistics

Table 2 presents the descriptive statistics for the variables used in my analysis, including the mean, standard deviation (shown in parentheses), and the within-firm standard deviation (indicated by brackets). In Panel A, where I summarize the statistics for variables derived from social media, a significant portion of the messages, averaging 51.1%, are classified as neutral, with happy messages constituting 24.6%. Analyzing the distribution of emotion variables reveals a tendency towards positive skewness, indicating that investors might be more inclined to express optimism rather than pessimism on social media platforms. Fear and surprise emerge as the third and fourth most prevalent emotions, respectively, followed by disgust, sadness, and anger. Similar to the return variables, the within-firm standard deviation for the emotion variables is only marginally lower than the overall sample standard deviation, suggesting substantial variability that can be utilized even when applying a firm-date fixed effect model. I plot the distribution of emotions in Panel (e) of Figure 1.

To provide context for the scale of my other variables, I compare my final estimation

dataset with a parallel sample from CRSP/Compustat for the same period, adhering to identical sample criteria (i.e., active, common ordinary shares on U.S. exchanges). The findings, presented in Panel B of Table 2, reveal several notable trends. First, it appears that StockTwits users are more inclined to talk about stocks that have recently increased in value or are currently on the rise: the mean past monthly return in my dataset is 3.3 percentage points higher, the close-to-open return is up by 0.3 percentage point, and the one-day lagged open-to-close return is 0.1 percentage point above those in the CRSP dataset. Secondly, these discussed stocks exhibit a 0.3 percentage point lower open-to-close return on average, hinting at potential mean-reversion. Additionally, I observe a heightened interest among social media investors in firms characterized by larger trading volumes, higher volatility, greater market capitalization, and increased short interest, but with lesser institutional ownership and lower historical returns.

Table 3 showcases the pairwise correlation coefficients for the variables in my analysis. These include my emotion variables, daily returns, and key control variables such as the lagged open-close return, close-open return, trading return over the past 20 days, and volatility over the past 183 days. The emotion measures show statistically significant correlations with daily returns and among themselves. Specifically, the neutral emotion category exhibits the highest correlations with both positively and negatively valenced emotions, serving as the baseline group, and hence omitted in the regression specifications. Additionally, to address the issue of a high correlation between certain negative emotions, such as anger and disgust, these can be merged into a single “hate” category, though this adjustment does not markedly influence the findings. The primary focus remains on the emotions of happiness and fear, which are crucial for understanding the market dynamics in terms of the classic dichotomy of fear and greed. I considered incorporating additional control variables, but variables like trading volume and market capitalization were highly correlated with volatility, leading to their exclusion to prevent multicollinearity. Although short interest and institutional ownership were not included in the main models due to their potential to significantly reduce

the sample size, I examine these factors separately in Table 8.

Table 1: Itemized Sample Restrictions

	Messages
StockTwits Data 2010-2021*	242,362,477
Keep	
Single Ticker	180,298,172
Not Automated	167,040,745
Active, Common Ordinary Shares, Traded on US Exchanges**	103,248,233
At Least 10 Messages during Non-Market/Market Hours	88,055,174
Final Pre-Market (4pm-9am) Sample	37,657,214
Final Market (9am-4pm) Sample	50,397,960

Notes: *Sample starts January 1st 2010 and runs till September 30th 2021. **I filter out stocks for which the security status (SECSTAT is "I") is inactive and restrict my sample to common ordinary shares (TPCI is "0") traded on US exchanges (EXCHG is 11, 12, 14, or 17).

Table 2: Summary Statistics

	CRSP/Compustat-StockTwits	CRSP/Compustat	Difference
Panel A: Social Media Information			
Happy	0.246 (0.150) [0.108]		
Sad	0.036 (0.0294) [0.0257]		
Fear	0.085 (0.0532) [0.0432]		
Disgust	0.037 (0.0343) [0.0284]		
Anger	0.024 (0.0213) [0.0174]		
Surprise	0.062 (0.0511) [0.0438]		
Neutral	0.511 (0.248) [0.155]		
Panel B: Financial Information			
Open-Close Return	-0.002 (0.0508) [0.0481]	0.000 (0.032)	-0.003***
Close-Open Return	0.004 (0.0410) [0.0390]	0.001 (0.022)	0.003***
Open-Close Return ₋₁	0.001 (0.0536) [0.0507]	0.000 (0.032)	0.001***
Return _{-20,-1}	0.047 (0.297) [0.267]	0.014 (0.158)	0.033***
\$ Volume _{-183,-1}	17.032 (2.615) [0.986]	15.229 (2.764)	1.803***
Volatility _{-183,-1}	0.039 (0.0217) [0.00932]	0.027 (0.017)	0.012***
Market Cap ₋₁	21.398 (2.624) [0.921]	20.742 (2.076)	0.655***
Institutional Ownership	0.518 (0.322) [0.106]	0.602 (0.318)	-0.084***
Short Interest	0.074 (0.0819) [0.0478]	0.042 (0.053)	0.032***
Observations	479,463	10,428,859	

Notes: \$ Volume and Market capitalization are in logs. Institutional Ownership, Short Interest, Return, and Volatility are in units. Close-Open Return refers to the return between closing price at $t - 1$ to the opening price at t . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard deviations in parentheses, within-firm standard deviation in brackets for my CRSP/Compustat-StockTwits sample only (I remove firm and date fixed effects to help interpret effect sizes).

Table 3: Correlation Matrix

	Daily Return	Happy	Sad	Fear	Disgust	Anger	Surprise	Neutral	Return _{oc,-1}	Return _{co}	Return _{-20,-1}
Happy	-0.03***										
Sad	-0.02***	0.26***									
Fear	-0.02***	0.33***	0.59***								
Disgust	-0.02***	0.30***	0.52***	0.48***							
Anger	-0.02***	0.34***	0.53***	0.51***	0.79***						
Surprise	-0.02***	0.32***	0.36***	0.38***	0.40***	0.42***					
Neutral	0.03***	-0.85***	-0.60***	-0.67***	-0.64***	-0.66***	-0.62***				
Return _{oc,-1}	-0.03***	0.11***	-0.12***	-0.10***	-0.11***	-0.10***	-0.05***	0.00			
Return _{co}	-0.07***	0.09***	-0.11***	-0.08***	-0.07***	-0.06***	-0.03***	-0.01*	0.03***		
Return _{-20,-1}	-0.02***	0.12***	-0.06***	-0.02***	-0.04***	-0.03***	-0.00	-0.05***	0.20***	-0.00	
Volatility _{-183,-1}	-0.04***	0.42***	0.17***	0.16***	0.26***	0.28***	0.24***	-0.42***	0.02***	0.06***	0.16***
Observations	479463										

Notes: Pairwise correlation with Bonferroni corrections. Continuous variables winsorized at the 0.1% and 99.9% levels. t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Return_{co} refers to return between the closing price at $t - 1$ and the opening price at t . Return_{oc} refers to the open-close return.

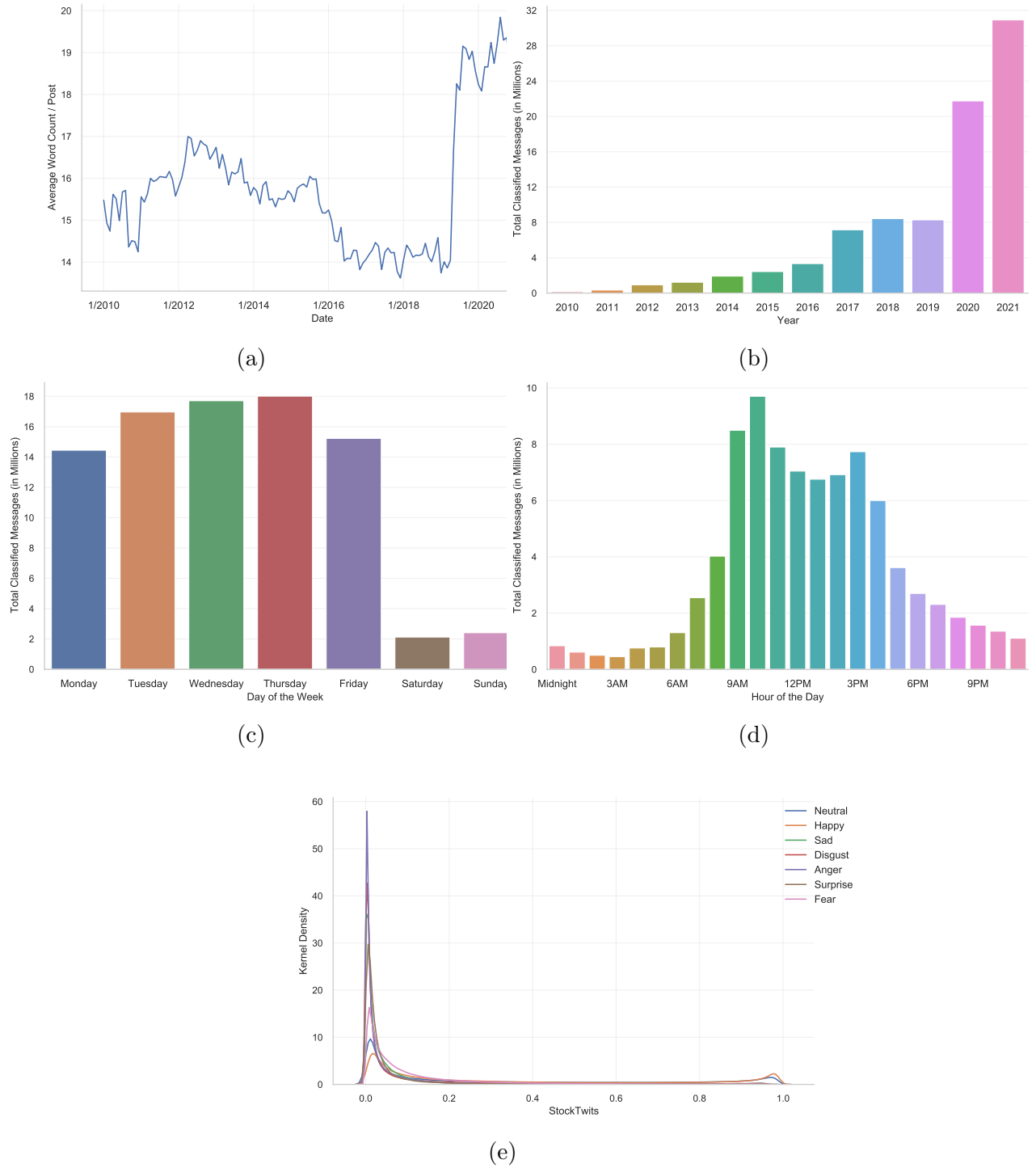


Figure 1: Post Characteristics

Notes: This includes my final dataset for pre-market and market sessions. In Panel (a), I chart the average word count per message over time, and Panel (b) illustrates the trend in the volume of posts in my dataset. Panel (c) presents the distribution of posts by day of the week, Panel (d) shows the distribution by hour of the day, and Panel (e) plots the distribution of classified messages.

3 Primary Findings

As demonstrated by Vamossy and Skog [2023], emotions can vary considerably over time. In this section, I delve into how such emotional shifts impact stock returns. I hypothesize that heightened investor enthusiasm could lead to increased buying activity, temporarily elevating stock prices. I explore this effect on daily price movements.

Before delving into the main findings, I must clarify that my analysis does not establish a causal relationship between emotions on social media and stock price changes. I demonstrate that emotions extracted from social media offer insights into stock valuation that are not covered by static stock characteristics, temporal patterns, or recent price trends. Moreover, this approach supports the insights gained from controlled lab experiments regarding the influence of investor emotions on trading behaviors.

3.1 Empirical Framework

My return regressions exploit within-firm variation. That is, I estimate the following model:

$$Y_{it} = \alpha + \sum_{j=0}^{j=5} \beta_j \text{Emotion}_{jit} + \gamma Y_{i,t-1} + \zeta X_{i,t-1} + \delta_t + \delta_i + \epsilon_{it} \quad (1)$$

where Emotion_{jit} represents the emotion metric for firm i on date t , calculated from messages posted during non-market hours before the market opens. In the regression framework, neutral emotions serve as the reference category, with j encompassing emotions such as happy, sad, anger, disgust, surprise, and fear. The primary dependent variable in most of my analysis is the daily open-close return. Day fixed effects (δ_t) are included to account for significant daily news events or overall market returns, while firm fixed effects (δ_i) help control for inherent characteristics of firms that might influence results, such as a firm's tendency to yield high returns possibly eliciting more excitement among investors before market opening. To address autocorrelation, I control for the previous day's open-close return. Additionally, I incorporate the close-open return (from the previous day's close to

the current day’s open) to verify that the emotions captured offer new information beyond after-hours and pre-market price changes. My models also factor in controls ($X_{i,t-1}$) for the return over the past 20 trading days and volatility. To dispel doubts that the emotion measure merely reflects shifts in trading volume, I base my dependent variable, the day t open-close return, on emotions from messages posted before the market opened on day t , ensuring clear temporal precedence of the emotion metric over the dependent variable.

In further refining my analysis, I categorize messages into two groups and calculate the emotion score for each. The first group consists of messages that directly reference a company’s fundamentals or stock price, which I label as “finance” messages. The second group, termed “chat” messages, encompasses more general discussions. This distinction aids in shedding light on the influence of emotions within varied informational contexts. Additionally, I differentiate messages based on their content: “original” messages are those that provide new information, characterized by not being a retweet and lacking hyperlinks, while “dissemination” messages are those that circulate existing information. This classification helps in understanding the impact of emotions in messages offering new insights versus those that simply share what is already known.

3.2 Testing Laboratory Experiments

I first delve into whether the outcomes observed in controlled laboratory experiments about investor emotions and their trading behaviors are mirrored in the investor emotion data I have collected. Table 4 categorizes the findings of pivotal laboratory experiments into directly and indirectly testable outcomes. To refine the reliability of these insights, I demean each metric by firm and date, aside from Finding III. By doing so, I close off certain mechanisms that could otherwise obscure the true relationship between emotions and market behavior.

In Panel A, I confirm the negative correlation between pre-market fear and subsequent price movements, in line with Breaban and Noussair [2018]. This is further substantiated by the observed correlations of fear, anger, and happiness with contemporaneous price move-

ments, echoing the findings of both Hargreaves Heap and Zizzo [2011] and Breaban and Noussair [2018]. Next, I extend the comparison to indirectly testable hypotheses, as delineated in Panel B of Table 4. To test these, I operate under two key assumptions: first, that traders exhibiting loss aversion tend to steer clear of firms with smaller market capitalizations, possibly due to perceived higher risks; and second, that there is a relationship between trading volume and cash holdings, where increased trading activity suggests lower cash reserves among traders. My findings in Panel B replicate the findings of lab experiments.

Despite noting somewhat modest correlation coefficients, I validated the conclusions of two laboratory experiments that link emotions to trading actions. This comparison not only reinforces the relevance of emotions in trading decisions but also highlights the potential of social media data to reflect complex emotional dynamics observed in more controlled experimental settings.

3.3 Emotions, Information Content, and Returns

I next explore the connection between investor emotions and daily stock returns through my regression framework specified in 1. Closest to my setting, lab evidence finds a positive (negative) relationship between investor enthusiasm (fear) before the market opens and subsequent returns (Breaban and Noussair [2018]). I empirically test this relationship in Table 5. In Column (1), I incorporate only my control variables, along with firm and date fixed effects. By directly comparing Column (1) to Column (2), I observe that emotions prior to market opening explain a small portion of the variance in daily returns, with an increase in R^2 by 0.05 percentage points. Additionally, I find that a one standard deviation increase in enthusiasm before the market opens on day t correlates with a 0.7% increase in the standard deviation of daily stock returns. In Column (3), my results indicate that the influence of investor emotions diminishes for larger firms and those that are easier to value, such as those within the S&P Super Composite index. My findings in Column (4) reveal more significant effects for stocks with higher user engagement (those with at least 100 mes-

sages), where a one standard deviation increase in pre-market happiness is associated with a 3.1% increase in the standard deviation of daily open-close stock returns. These insights suggest that investor emotions, as derived from StockTwits, provide valuable information relevant to stock valuation that is not accounted for by time-invariant stock characteristics, by temporal patterns, or by recent price movements.

Table 6 repeats my analysis using measures of emotions disaggregated between messages that convey earnings or trade-related information (“finance”) and those that provide other types of information (“chat”) in Column (1) and Column (2). I observe that emotions derived from messages specifically mentioning firm fundamentals and earnings generally present smaller point estimates for emotions with negative valence. This indicates that seemingly irrelevant messages carry important information that I capture by extracting emotions. For example, a user might post “:) :)” in response to news events or other posts, which clearly express their excitement but do not include any direct information about the stock beyond their emotional state. Furthermore, when comparing Column (3) to Column (4), I note that both messages containing original information and those disseminating existing information contribute to my findings. Notably, in the context of messages that disseminate existing information, only the levels of fear and happiness show statistically significant deviations from the baseline (neutral) level.

3.4 Emotions and Persistence

Table 7 explores the impact on emotions before the market opens today on returns the following four days. As expected, I find that the predictive power of investor emotions diminishes over time. Interestingly, I do find persistent negative impacts for sadness, suggesting its uniquely enduring influence on market dynamics. Furthermore, in Panel B, when I consolidate these emotional measures into a singular metric, valence⁴, I observe a reduction in both

⁴I follow Breaban and Noussair [2018] and define valence as:

$$\text{Valence} = \text{Happy} - \text{Sad} - \text{Anger} - \text{Disgust} - \text{Fear} \quad (2)$$

the predictive power (evidenced by lower R^2 values) and the statistical significance of the emotions. This outcome highlights the inherent value in examining granular emotions rather than relying on a one-dimensional measure like valence.

Table 4: Empirical tests of Lab Experiments on Emotions and Asset Pricing

Finding	Source	Dependent Variable	Sample	Test	Result	
Panel A: Directly Testable						
<i>Finding I:</i> Average trader fear before the market opens is negatively correlated with subsequent price movements.	Breaban and Noussair [2018]	Return	Pre-market	$\rho_{y,fear} < 0$	-0.0045^{***}	✓
<i>Finding II:</i> Fear and anger are negatively, while happiness is positively correlated with price movements.	Hargreaves Heap and Zizzo [2011], Breaban and Noussair [2018]	Return	Market	$\rho_{y,fear} < 0$, $\rho_{y,anger} < 0$, $\rho_{y,happy} > 0$	-0.1879^{***} , -0.1684^{***} , 0.2337^{***}	✓ ✓ ✓
Panel B: Indirectly Testable						
<i>Finding III:</i> Traders with higher levels of fear (valence) make more (less) loss-averse decisions. <i>Alternative III:</i> Average trader fear (valence) before the market opens is positively (negatively) correlated with the ease of valuation.	Breaban and Noussair [2018]	Market Cap	Pre-market	$\rho_{y,fear} > 0$, $\rho_{y,valence} < 0$	0.0151^{***} , — 0.1717^{***}	✓ ✓
<i>Finding IV:</i> Cash holdings are negatively correlated with fear. <i>Alternative IV:</i> Trading volume is positively correlated with fear.	Breaban and Noussair [2018]	Volume	Market	$\rho_{y,fear} > 0$	0.0161^{***}	✓

Notes: This table investigates whether the findings of controlled laboratory experiments that relate investor emotions to trading behavior replicate my observational data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level. Aside from Finding III, variables are demeaned by firm and date.

Table 5: Pre-Market Emotions and Price Movements

	(1)	(2)	(3)	(4)
Happy		0.0052*** (0.0007)	0.0007 (0.0008)	0.0430*** (0.0064)
Sad		-0.0242*** (0.0031)	-0.0096*** (0.0037)	-0.0341 (0.0279)
Fear		-0.0064*** (0.0018)	-0.0047** (0.0019)	0.0094 (0.0156)
Disgust		-0.0041 (0.0033)	-0.0042 (0.0042)	-0.0233 (0.0239)
Anger		-0.0050 (0.0055)	-0.0036 (0.0074)	-0.0062 (0.0423)
Surprise		-0.0068*** (0.0017)	-0.0057*** (0.0021)	0.0267* (0.0148)
S&P Super Composite Firms			X	
At least 100 messages				X
$\sigma_{y,within}$	0.0475	0.0475	0.0308	0.0669
Observations	454,328	454,328	163,011	53,803
R^2	0.1047	0.1052	0.1443	0.1742

Notes: This table considers the relationship between pre-market emotions and daily price movements. The dependent variable is daily returns, computed as the difference between the closing and the opening price, divided by the opening price. All specifications include firm and date fixed effects, close-open returns, lag open-close returns, past 20 trading days return, and volatility (as defined in Variable Definitions). Robust standard errors clustered at the industry and date levels are in parentheses. I use the Fama-French 12-industry classification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level to mitigate the impact of outliers. I report the within-firm, detrended (demeaned) standard deviation of the dependent variable.

Table 6: Pre-Market Emotions, Message Content and Price Movements

	(1)	(2)	(3)	(4)
	<u>Chat Type</u>		<u>Information Type</u>	
	Chat	Finance	Disseminating	Original
Happy	0.0028*** (0.0004)	0.0025*** (0.0006)	0.0030*** (0.0005)	0.0027*** (0.0005)
Sad	-0.0063*** (0.0016)	-0.0155*** (0.0025)	-0.0045** (0.0020)	-0.0104*** (0.0018)
Fear	-0.0021* (0.0012)	-0.0039*** (0.0014)	-0.0016 (0.0012)	-0.0043*** (0.0011)
Disgust	-0.0037** (0.0016)	-0.0020 (0.0028)	-0.0023 (0.0022)	-0.0047** (0.0019)
Anger	-0.0010 (0.0026)	-0.0061 (0.0044)	-0.0044 (0.0035)	-0.0041 (0.0029)
Surprise	-0.0031*** (0.0008)	-0.0027** (0.0013)	0.0005 (0.0010)	-0.0037*** (0.0009)
$\sigma_{y,within}$	0.0484	0.0475	0.0475	0.0487
Observations	419,232	454,153	448,054	413,359
R^2	0.1078	0.1050	0.1052	0.1098

Notes: This table considers the relationship between pre-market emotions disaggregated by information content and information type and daily price movements. The dependent variable is daily returns, computed as the difference between the closing and the opening price, divided by the opening price. All specifications include firm and date fixed effects, close-open returns, lag open-close returns, past 20 trading days return, and volatility (as defined in Variable Definitions). Robust standard errors clustered at the industry and date levels are in parentheses. I use the Fama-French 12-industry classification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level to mitigate the impact of outliers. I report the within-firm, detrended (demeaned) standard deviation of the dependent variable.

Table 7: Pre-Market Emotions and Subsequent Price Movements

	(1) Ret _{t+1}	(2) Ret _{t+2}	(3) Ret _{t+3}	(4) Ret _{t+4}
Panel A: Granular Emotions				
Happy	0.0004 (0.0007)	0.0002 (0.0007)	-0.0003 (0.0006)	0.0005 (0.0006)
Sad	-0.0090*** (0.0029)	-0.0060** (0.0028)	-0.0050* (0.0028)	-0.0021 (0.0028)
Fear	0.0026 (0.0017)	0.0010 (0.0017)	0.0039** (0.0017)	-0.0002 (0.0016)
Disgust	0.0069** (0.0033)	0.0037 (0.0032)	-0.0036 (0.0032)	0.0062* (0.0032)
Anger	0.0025 (0.0055)	0.0034 (0.0055)	0.0096* (0.0055)	0.0024 (0.0053)
Surprise	-0.0040** (0.0016)	-0.0010 (0.0015)	-0.0025 (0.0015)	-0.0010 (0.0015)
R^2	0.1113	0.1141	0.1180	0.1160
Panel B: Combined Emotions				
Valence	-0.0002 (0.0005)	-0.0001 (0.0005)	-0.0004 (0.0005)	-0.0002 (0.0005)
R^2	0.1107	0.1140	0.1180	0.1160
$\sigma_{y,within}$	0.0431	0.0418	0.0410	0.0405
Observations	454,169	454,067	453,986	453,878

Notes: This table investigates the predictive power of investor emotions on price movements one to four days into the future. The dependent variable is daily returns, computed as the difference between the closing and the opening price, divided by the opening price at $t+1$ through $t+4$. All specifications include firm and date fixed effects, close-open returns, open-close returns at $t-1$, past 20 trading days return and volatility (as defined in Variable Definitions). Robust standard errors clustered at the industry and date levels are in parentheses. I use the Fama-French 12-industry classification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level to mitigate the impact of outliers. I report the within-firm, detrended (demeaned) standard deviation of the dependent variable.

4 Additional Findings

The findings up to this point indicate that investor emotions, as extracted from social media, offer insightful information capable of forecasting daily stock price fluctuations. Nevertheless, the impact of these emotions is unlikely to be uniform. In this section, I delve into the heterogeneous effects linked to both stock and user characteristics, demonstrating that my results are not only robust but also intuitively reflect the underlying investor emotions.

4.1 Heterogeneous Effects: Stock Characteristic

The theoretical framework surrounding investor sentiment suggests that stocks that are younger, smaller, more volatile, unprofitable, do not pay dividends, and are in distress, exhibit a higher sensitivity to investor sentiment. On the other hand, stocks that resemble bonds in their characteristics tend to be less influenced by sentiment fluctuations, as discussed in Baker and Wurgler [2007].

4.1.1 Emotions, Volatility, and Liquidity

To investigate if emotions mirror the behavior of sentiment, I interact my emotion variables with a dummy variable designed to encapsulate volatility and liquidity aspects, such as market capitalization and dollar trading volume. Consistent with theoretical insights on investor sentiment, I find larger point estimates for investor enthusiasm in firms characterized by smaller market capitalization (as seen in Column (4) of Table 8).

4.1.2 Emotions, Short Interest and Institutional Ownership

I next interact my emotion variables with institutional ownership and short interest. I find larger impacts when short interest is higher (Column (5) of Table 8).

4.2 Heterogeneous Effects: User Characteristic

Hong and Page [2004] demonstrate that a diverse group of intelligent decision-makers consistently outperforms a more homogenous group of highly skilled individuals. I explore this concept by categorizing the messages I have collected based on the traders' investment horizons (i.e., long-term vs. short-term), trading methodologies (i.e., value vs. technical), and levels of trading experience (i.e., amateur vs. professional).

In Table 9, I report the heterogeneity across user types and make several notable observations. First, consistent with the hypothesis on the benefits of diversity, I discover that the emotions expressed by homogeneous groups offer less predictive power regarding daily price fluctuations. Second, I observe that the correlation between investor emotions and daily stock movements remains consistent and statistically significant across the majority of my specifications.

4.3 Sensitivity Analysis

Next, I show that my results are robust and capture investor emotions intuitively.

4.3.1 Alternative Classification

I use alternative emotion variables using a model trained on Twitter emotion metadata collected by other researchers. I examine this Twitter-based model in Columns (1-3) of Table 10. My results continue to hold significance, lending robust support to the link between investor emotions and daily stock movements. Vamossy and Skog [2023] provide a detailed explanation for the preference for the StockTwits model over the Twitter model.

4.3.2 Alternative Weighting

For my previous findings, I applied a weighting to messages based on $\log(1+\text{followers})$. As a measure to test the validity of this approach, I explore a different weighting scheme. Specifically, I consider a scenario where I discard the initial weighting approach, thereby

treating all messages with equal importance (as shown in Columns (4-6) of Table 10). My results remain mostly consistent even with these alternative specifications.

4.3.3 Contemporaneous Emotions & Prices

To further validate that my emotion metrics accurately reflect investor emotions, I also run my primary regression using contemporaneous (within trading hours) emotion variables. Incorporating emotions measured during market hours into the analysis considerably enhances the model’s robustness. The results presented in Table 11 reinforce the accuracy of the happiness metric in capturing investor enthusiasm. Specifically, by including contemporaneous emotion variables, the R^2 of the primary regression experienced a significant increase of 8.46 percentage points, which constitutes a 75% improvement in the model’s explanatory power. This substantial improvement in R^2 underscores the added value of considering investor emotions within trading hours, affirming the strong link between positive market returns and elevated levels of enthusiasm among investors. Furthermore, the point estimates for these emotion variables increased by more than an order of magnitude, which supports the argument that these emotion metrics serve as reliable, multi-dimensional, and granular indicators of investor sentiment, especially during active trading periods. Nonetheless, I refrain from delving deeper into the contemporaneous effects, as they might be reactive rather than predictive.

Table 8: Pre-Market Emotions, Stock Characteristics and Price Movements

	(1)	(2)	(3)	(4)	(5)	(6)
	None	\$ Volume _{p25}	Volatility _{p75}	Interaction Variable Market Cap _{p25}	Short Interest _{p75}	Institutional _{p25}
Happy	0.0048*** (0.0007)	0.0057*** (0.0007)	0.0043*** (0.0007)	0.0059*** (0.0007)	0.0040*** (0.0008)	0.0040*** (0.0011)
Sad	-0.0238*** (0.0031)	-0.0175*** (0.0032)	-0.0232*** (0.0031)	-0.0186*** (0.0032)	-0.0224*** (0.0036)	-0.0267*** (0.0044)
Anger	-0.0068 (0.0055)	0.0044 (0.0059)	-0.0048 (0.0054)	-0.0061 (0.0059)	-0.0080 (0.0064)	-0.0129 (0.0081)
Surprise	-0.0070*** (0.0017)	-0.0065*** (0.0018)	-0.0052*** (0.0017)	-0.0060*** (0.0017)	-0.0059*** (0.0020)	-0.0061** (0.0025)
Disgust	-0.0029 (0.0033)	-0.0110*** (0.0034)	-0.0039 (0.0032)	-0.0068** (0.0034)	0.0017 (0.0040)	-0.0036 (0.0048)
Fear	-0.0070*** (0.0018)	-0.0050*** (0.0018)	-0.0065*** (0.0018)	-0.0060*** (0.0018)	-0.0064*** (0.0021)	-0.0090*** (0.0025)
Happy $\times$ IV		-0.0020 (0.0016)	0.0022 (0.0018)	-0.0032* (0.0017)	0.0027* (0.0014)	-0.0006 (0.0027)
Sad $\times$ IV		-0.0250*** (0.0081)	-0.0044 (0.0088)	-0.0190** (0.0079)	-0.0054 (0.0065)	-0.0171 (0.0110)
Anger $\times$ IV		-0.0294** (0.0132)	-0.0015 (0.0144)	-0.0013 (0.0124)	0.0018 (0.0114)	0.0165 (0.0202)
Surprise $\times$ IV		-0.0011 (0.0042)	-0.0064 (0.0046)	-0.0031 (0.0040)	-0.0048 (0.0034)	0.0038 (0.0061)
Disgust $\times$ IV		0.0217*** (0.0080)	-0.0008 (0.0089)	0.0090 (0.0076)	-0.0145** (0.0069)	0.0009 (0.0114)
Fear $\times$ IV		-0.0065 (0.0049)	-0.0011 (0.0054)	-0.0037 (0.0048)	-0.0033 (0.0038)	0.0118* (0.0071)
IV		0.0046*** (0.0007)	0.0001 (0.0008)	0.0072*** (0.0008)	0.0020*** (0.0006)	-0.0011 (0.0013)
$\sigma_{y,within}$	0.0477	0.0475	0.0475	0.0477	0.0476	0.0451
Observations	467,506	454,364	454,328	467,506	467,196	189,588
R ²	0.1053	0.1053	0.1052	0.1061	0.1055	0.1096

Notes: This table considers the relationship between pre-market emotions, stock characteristics, and daily price movements. The dependent variable is daily returns, computed as the difference between the closing and the opening price, divided by the opening price. The interaction variables (IV) are dummy variables for \$ trading Volume (Column (2)), volatility (Column (3)), market cap (Column (4)), short interest (Column (5)), and institutional ownership (Column (6)). All specifications include firm and date fixed effects, close-open returns, lag open-close returns, and past 20 trading days returns. Robust standard errors clustered at the industry and date levels are in parentheses. I use the Fama-French 12-industry classification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level to mitigate the impact of outliers. I report the within-firm, detrended (demeaned) standard deviation of the dependent variable.

Table 9: Pre-Market Emotions, Investor Types and Price Movements

	(1)	(2)	(3)	(4)	(5)	(6)
	Trading Experience		Trading Approach		Investment Horizon	
	Amateur	Professional	Fundamental	Technical	Short-Term	Long-Term
Happy	0.0010** (0.0004)	0.0009** (0.0004)	0.0018*** (0.0004)	0.0012*** (0.0004)	0.0018*** (0.0004)	0.0016*** (0.0004)
Sad	-0.0064*** (0.0014)	-0.0041** (0.0016)	-0.0029** (0.0014)	-0.0068*** (0.0014)	-0.0058*** (0.0013)	-0.0052*** (0.0015)
Anger	-0.0030 (0.0024)	-0.0051 (0.0031)	-0.0039 (0.0025)	-0.0065** (0.0026)	-0.0042* (0.0025)	-0.0009 (0.0026)
Surprise	-0.0026*** (0.0008)	-0.0018** (0.0009)	-0.0012 (0.0008)	-0.0019** (0.0009)	-0.0011 (0.0008)	-0.0009 (0.0008)
Disgust	-0.0042*** (0.0015)	-0.0006 (0.0020)	-0.0001 (0.0016)	-0.0027* (0.0016)	-0.0032** (0.0016)	-0.0035** (0.0017)
Fear	-0.0021** (0.0009)	-0.0022** (0.0010)	-0.0028*** (0.0009)	-0.0007 (0.0009)	-0.0007 (0.0009)	-0.0018* (0.0009)
Observations	369,668	385,533	378,307	385,460	397,150	410,853
R^2	0.1126	0.1051	0.1063	0.1058	0.1046	0.1071

Notes: This table considers the relationship between pre-market emotions, user characteristics, and daily price movements. The dependent variable is daily returns, computed as the difference between the closing and the opening price, divided by the opening price. All specifications include firm and date fixed effects, close-open returns, lag open-close returns, past 20 trading days return, and volatility (as defined in Variable Definitions). Robust standard errors clustered at the industry and date levels are in parentheses. I use the Fama-French 12-industry classification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level to mitigate the impact of outliers. I report the within-firm, detrended (demeaned) standard deviation of the dependent variable.

Table 10: Robustness: Pre-Market Emotions and Price Movements

	(1)	(2)	(3)	(4)	(5)	(6)
	Twitter Model			Alternative Weighting		
Happy	0.0025*** (0.0009)	0.0011 (0.0010)	0.0086 (0.0103)	0.0055*** (0.0007)	0.0006 (0.0008)	0.0497*** (0.0075)
Sad	-0.0207*** (0.0017)	-0.0084*** (0.0017)	-0.0602*** (0.0215)	-0.0266*** (0.0031)	-0.0077** (0.0036)	-0.0361 (0.0286)
Anger	-0.0116*** (0.0023)	-0.0063** (0.0029)	-0.0265 (0.0179)	-0.0088 (0.0055)	-0.0069 (0.0071)	-0.0243 (0.0430)
Surprise	0.0011 (0.0015)	0.0009 (0.0016)	-0.0017 (0.0202)	-0.0086*** (0.0017)	-0.0057*** (0.0020)	0.0411*** (0.0157)
Disgust	0.0131 (0.0158)	0.0003 (0.0156)	-0.1245 (0.1272)	-0.0028 (0.0033)	-0.0033 (0.0040)	-0.0081 (0.0243)
Fear	-0.0040** (0.0016)	-0.0055*** (0.0017)	0.0267 (0.0163)	-0.0068*** (0.0019)	-0.0049** (0.0020)	0.0183 (0.0168)
S&P Super Composite Firms	X			X		
At least 100 messages	X			X		
$\sigma_{y,within}$	0.0481	0.0309	0.0673	0.0481	0.0309	0.0673
Observations	454,328	163,011	53,803	454,328	163,011	53,803
R^2	0.1052	0.1444	0.1735	0.1053	0.1443	0.1743

Notes: This table considers the robustness of the relationship between pre-market emotions and daily price movements. Columns (1-3) use emotion variables constructed based on a model trained on Twitter emotion data compiled by other researchers. Columns (4-6) abandon the weighting scheme used in my analysis and weigh each posts equally. The dependent variable is daily returns, computed as the difference between the closing and the opening price, divided by the opening price. All specifications include firm and date fixed effects, close-open returns, lag open-close returns, past 20 trading days return, and volatility (as defined in Variable Definitions). Robust standard errors clustered at the industry and date levels are in parentheses. I use the Fama-French 12-industry classification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level to mitigate the impact of outliers. I report the within-firm, detrended (demeaned) standard deviation of the dependent variable.

Table 11: Market Emotions and Price Movements

	(1)	(2)	(3)	(4)
Happy		0.0821*** (0.0008)	0.0547*** (0.0008)	0.2535*** (0.0064)
Sad		-0.1892*** (0.0027)	-0.1102*** (0.0033)	-0.8765*** (0.0220)
Fear		-0.1003*** (0.0018)	-0.0507*** (0.0020)	-0.4925*** (0.0145)
Surprise		-0.0619*** (0.0017)	-0.0456*** (0.0021)	-0.0777*** (0.0141)
Disgust		-0.1173*** (0.0029)	-0.0500*** (0.0036)	-0.3808*** (0.0214)
Anger		-0.0677*** (0.0044)	-0.0354*** (0.0058)	-0.0539* (0.0321)
S&P Super Composite Firms			X	
At least 100 messages				X
$\sigma_{y,within}$	0.0539	0.0539	0.0350	0.0773
Observations	476,802	476,801	160,354	78,268
R^2	0.1129	0.1975	0.2292	0.3389

Notes: This table considers the relationship between emotions posted during market hours and daily price movements. The dependent variable is daily returns, computed as the difference between the closing and the opening price, divided by the opening price. All specifications include firm and date fixed effects, close-open returns, lag open-close returns, past 20 trading days return, and volatility (as defined in Variable Definitions). Robust standard errors clustered at the industry and date levels are in parentheses. I use the Fama-French 12-industry classification. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Continuous variables winsorized at the 0.1% and 99.9% level to mitigate the impact of outliers. I report the within-firm, detrended (demeaned) standard deviation of the dependent variable.

5 Conclusion

This work contributes to the literature on investor emotions and market dynamics, corroborating the results from controlled laboratory experiments. I have shown that investor emotions are significant predictors of daily stock price movements, with their influence being more pronounced when accompanied by greater short interest, and lower levels of liquidity. Furthermore, the persistence of sadness on subsequent returns and the insignificance of the combined valence metric underscores the importance of dissecting emotional states to gain a deeper and more accurate insight into how investor sentiments drive market movements.

References

- Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.
- Jeffrey S Abarbanell. Do analysts’ earnings forecasts incorporate information in prior stock price changes? *Journal of Accounting and Economics*, 1991.
- George A Akerlof and Robert J Shiller. *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton university press, 2010.
- Stefania Albanesi and Domonkos F Vamossy. Predicting consumer default: A deep learning approach. Technical report, National Bureau of Economic Research, 2019.
- Eduardo B Andrade, Terrance Odean, and Shengle Lin. Bubbling with excitement: an experiment. *Review of Finance*, 20(2):447–466, 2016.
- Werner Antweiler and Murray Z Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294, 2004.
- David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006.
- Susan Athey and Guido W Imbens. Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 2019.
- Pablo D Azar and Andrew W Lo. The wisdom of twitter crowds: Predicting stock market reactions to fomic meetings via twitter feeds. *The Journal of Portfolio Management*, 42(5):123–134, 2016.
- Malcolm Baker and Jeffrey Wurgler. Investor sentiment in the stock market. *Journal of economic perspectives*, 21(2):129–152, 2007.

- Brad M Barber and Terrance Odean. The behavior of individual investors. In *Handbook of the Economics of Finance*, volume 2, pages 1533–1570. Elsevier, 2013.
- Eli Bartov, Lucile Faurel, and Partha S Mohanram. Can twitter help predict firm-level earnings and stock returns? *The Accounting Review*, 93(3):25–57, 2018.
- Joyce Berg, Robert Forsythe, and Thomas Rietz. What makes markets predict well? evidence from the iowa electronic markets. In *Understanding Strategic Interaction*, pages 444–463. Springer, 1997.
- Joyce Berg, Robert Forsythe, Forrest Nelson, and Thomas Rietz. Results from a dozen years of election futures markets research. *Handbook of experimental economics results*, 1: 742–751, 2008.
- Elizabeth Blankespoor, Gregory S Miller, and Hal D White. The role of dissemination in market liquidity: Evidence from firms’ use of twitterTM. *The Accounting Review*, 89(1): 79–112, 2014.
- Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.
- Laura Ana Maria Bostan and Roman Klinger. An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119. Association for Computational Linguistics, 2018. URL <http://aclweb.org/anthology/C18-1179>.
- Adriana Breaban and Charles N Noussair. Emotional state and market behavior. *Review of Finance*, 22(1):279–309, 2018.
- Cristian Buciluă, Rich Caruana, and Alexandru Niculescu-Mizil. Model compression. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 535–541, 2006.
- Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.
- Hailiang Chen, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5):1367–1403, 2014.
- François Chollet et al. Keras: Deep learning library for theano and tensorflow. URL: <https://keras.io/k>, 7(8), 2015.
- Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- J Anthony Cookson and Marina Niessner. Why don’t we agree? evidence from a social network of investors. *The Journal of Finance*, 75(1):173–228, 2020.

- Asher Curtis, Vernon J Richardson, and Roy Schmardebeck. Investor attention and the pricing of earnings news. *Handbook of Sentiment Analysis in Finance, Forthcoming*, 2014.
- Zhi Da, Joseph Engelberg, and Pengjie Gao. In search of attention. *The Journal of Finance*, 66(5):1461–1499, 2011.
- Sanjiv R Das and Mike Y Chen. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, 53(9):1375–1388, 2007.
- J Bradford De Long, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738, 1990.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- Michael S Drake, Darren T Roulstone, and Jacob R Thornock. Investor information demand: Evidence from google searches around earnings announcements. *Journal of Accounting research*, 50(4):1001–1040, 2012.
- Darren Duxbury, Tommy Gärling, Amelie Gamble, and Vian Klass. How emotions influence behavior in financial markets: a conceptual analysis and emotion-based account of buy-sell preferences. *The European Journal of Finance*, pages 1–22, 2020.
- Alex Edmans, Diego Garcia, and Øyvind Norli. Sports sentiment and stock returns. *The Journal of finance*, 62(4):1967–1998, 2007.
- Katrina Ellis, Roni Michaely, and Maureen O’hara. When the underwriter is the market maker: An examination of trading in the ipo aftermarket. *The Journal of Finance*, 55(3): 1039–1074, 2000.
- Larry G Epstein and Martin Schneider. Ambiguity, information quality, and asset pricing. *The Journal of Finance*, 63(1):197–228, 2008.
- Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*, 2017.
- Peter Gabrovšek, Darko Aleksovski, Igor Mozetič, and Miha Grčar. Twitter sentiment around the earnings announcement events. *PloS one*, 12(2), 2017.
- John Kenneth Galbraith. *A short history of financial euphoria*, volume 3856. Penguin Books, 1994.
- Eric Gilbert and Karrie Karahalios. Widespread worry and the stock market. In *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.
- Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005.

- Shaun P Hargreaves Heap and Daniel John Zizzo. Emotions and chat in a financial markets experiment. *Available at SSRN 1783462*, 2011.
- Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015.
- Mark Hirschey, Vernon J Richardson, and Susan Scholz. Stock-price effects of internet buy-sell recommendations: The motley fool case. *Financial Review*, 35(2):147–174, 2000.
- David Hirshleifer and Tyler Shumway. Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032, 2003.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Lu Hong and Scott E Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46):16385–16389, 2004.
- John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90–95, 2007.
- Ben Jacobsen and Wessel Marquering. Is it the weather? *Journal of Banking & Finance*, 32(4):526–540, 2008.
- Russell Jame, Rick Johnston, Stanimir Markov, and Michael C Wolfe. The value of crowd-sourced earnings forecasts. *Journal of Accounting Research*, 54(4):1077–1110, 2016.
- Narasimhan Jegadeesh and Woojin Kim. Do analysts herd? an analysis of recommendations and market reactions. *The Review of Financial Studies*, 23(2):901–937, 2010.
- Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91, 1993.
- Michael J Jung, James P Naughton, Ahmed Tahoun, and Clare Wang. Do firms strategically disseminate? evidence from corporate use of social media. *The Accounting Review*, 93(4):225–252, 2018.
- Mark J Kamstra, Lisa A Kramer, and Maurice D Levi. Winter blues: A sad stock market cycle. *American Economic Review*, 93(1):324–343, 2003.
- John Maynard Keynes. *General Theory of Employment, Interest, and Money*. Palgrave Macmillan, 1936.
- Yaron Lahav and Shireen Meer. The effect of induced mood on prices in experimental asset markets. *Unpublished working paper, Emory University*, 2010.
- Jessica LaVoice and Domonkos F Vamossy. Racial disparities in debt collection. *arXiv preprint arXiv:1910.02570*, 2019.

- Alastair Lawrence, James Ryans, Estelle Sun, and Nikolay Laptev. Yahoo finance search and earnings announcements. *Available at SSRN 2804353*, 2016.
- Lian Fen Lee, Amy P Hutton, and Susan Shu. The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research*, 53(2):367–404, 2015.
- Qian Li, Bing Zhou, and Qingzhong Liu. Can twitter posts predict stock behavior?: A study of stock market with twitter social emotion. In *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, pages 359–364. IEEE, 2016.
- Alexander Ljungqvist, Vikram Nanda, and Rajdeep Singh. Hot markets, investor sentiment, and ipo pricing. *the Journal of Business*, 79(4):1667–1702, 2006.
- George F Loewenstein, Elke U Weber, Christopher K Hsee, and Ned Welch. Risk as feelings. *Psychological bulletin*, 127(2):267, 2001.
- Cheng Long, Brian M Lucey, and Larisa Yarovaya. ” i just like the stock” versus” fear and loathing on main street”: The role of reddit sentiment in the gamestop short squeeze. *SSRN Electronic Journal*, page 31, 2021.
- Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.
- Tim Loughran and Jay Ritter. Why has ipo underpricing changed over time? *Financial management*, pages 5–37, 2004.
- Tim Loughran and Jay R Ritter. The new issues puzzle. *The Journal of finance*, 50(1): 23–51, 1995.
- Robert E Lucas Jr. Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, pages 1429–1445, 1978.
- Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017. URL <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- Yuxin Mao, Wei Wei, Bing Wang, and Benyuan Liu. Correlating s&p 500 stocks with twitter data. In *Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research*, pages 69–72, 2012.
- Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.
- Vitaliy Meursault. The language of earnings announcements. Technical report, Working paper, 2019.

- Sendhil Mullainathan and Jann Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.
- Vivek Narayanan, Ishan Arora, and Arjun Bhatia. Fast and accurate sentiment classification using an enhanced naive bayes model. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 194–201. Springer, 2013.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct): 2825–2830, 2011.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
- Mitchell A Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1):435–480, 2009.
- Alexander Ratner, Stephen H Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. Snorkel: Rapid training data creation with weak supervision. *The VLDB Journal*, 29(2):709–730, 2020.
- Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. <http://is.muni.cz/publication/884893/en>.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.
- Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61: 85–117, 2015.
- Mike Schuster and Kuldeep K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.
- Robert J Shiller. Bubbles, human judgment, and expert opinion. *Financial Analysts Journal*, 58(3):18–26, 2002.
- Robert J Shiller. From efficient markets theory to behavioral finance. *Journal of economic perspectives*, 17(1):83–104, 2003.
- Robert J. Shiller. 'gut feelings' are driving the markets. *The New York Times*, 2020. URL <https://www.nytimes.com/2020/01/02/business/gut-feelings-are-driving-the-markets.html>.

- Hui-Chu Shu. Investor mood and financial markets. *Journal of Economic Behavior & Organization*, 76(2):267–282, 2010.
- Vernon L Smith, Gerry L Suchanek, and Arlington W Williams. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica: Journal of the Econometric Society*, pages 1119–1151, 1988.
- Douglas E Stevens and Arlington W Williams. Inefficiency in earnings forecasts: Experimental evidence of reactions to positive vs. negative information. *Experimental Economics*, 7(1):75–92, 2004.
- James Surowiecki. The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*, 296, 2004.
- Lazaros Symeonidis, George Daskalakis, and Raphael N Markellos. Does the weather affect stock market volatility? *Finance Research Letters*, 7(4):214–223, 2010.
- Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168, 2007.
- Robert Tumarkin and Robert F Whitelaw. News or noise? internet postings and stock prices. *Financial Analysts Journal*, 57(3):41–51, 2001.
- Domonkos F. Vamossy. Investor emotions and earnings announcements. *Journal of Behavioral and Experimental Finance*, 30:100474, 2021. ISSN 2214-6350. doi: <https://doi.org/10.1016/j.jbef.2021.100474>. URL <https://www.sciencedirect.com/science/article/pii/S2214635021000186>.
- Domonkos F Vamossy. Social media emotions and ipo returns. *arXiv preprint arXiv:2306.12602*, 2023.
- Domonkos F Vamossy and Rolf Skog. Emtract: Extracting emotions from social media. *Available at SSRN 3975884*, 2023.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.
- Ivo Welch. Herding among security analysts. *Journal of Financial economics*, 58(3):369–396, 2000.
- Justin Wolfers. How emotion hurts stock returns. *The New York Times*, 2015. URL <https://www.nytimes.com/2015/08/25/upshot/how-emotion-hurts-stock-returns.html>.
- Michelle Yik, James A Russell, and James H Steiger. A 12-point circumplex structure of core affect. *Emotion*, 11(4):705, 2011.

Appendix

A Variable Definitions and Sources

Table A.1 defines the variables used in the analyses.

Table A.1: Variable Definitions

Variable	Definition	Source
Open-Close Return	Difference between daily closing and open price, normalized by the open price.	CRSP
Close-Open Return	Difference between the previous day's closing price and current open price, normalized by the previous day's closing price.	CRSP
Market Cap ₋₁	Natural logarithm of the market value of equity ($\log(1 + \text{CSHOC} * \text{PRCCD})$).	CRSP
Volatility	Standard deviation of daily returns from 183 days prior up to a day before.	CRSP
\$ Volume _{-183,-1}	Arithmetic mean of the natural logarithm of market value of daily trading volume ($\log(1 + \text{CSHTRD} * (\text{PRCCD} + \text{PRCOD})/2)$) from 183 days prior up to a day before.	CRSP
Short Interest	Shares short divided by shares outstanding. Bi-weekly frequency.	Compustat
Institutional Ownership	Number of shares held by institutional investors scaled by total shares outstanding as of the quarter-end date	Thomson Reuters Institutional Holdings (13F)
Emotion	Each message is classified by a many-to-one deep learning model into one of the seven categories (i.e., neutral, happy, sad, anger, disgust, surprise, fear) so that the corresponding probabilities sum up to 1. For each emotion separately, I then take the weighted average of these probabilities from 4 pm the previous day until 9.29 am the trading day. The weights correspond to the number of followers of the user $1 + \log(1 + \# \text{ of Followers})$.	StockTwits

B StockTwits Activity & Sample Distributions

Table B.1 displays the descriptive statistics of StockTwits activity for my dataset, particularly focusing on the annual frequency distributions. There is a noticeable escalation in

StockTwits usage throughout my study period, with the number of messages surging from 49,074 in 2010 to 10,584,254 by 2020. This trend underscores the burgeoning role of social media over the 2010s. Likewise, the scope of my analysis broadened significantly, capturing firm-day observations that increased from 1,660 in 2010 to 129,565 by 2020.

In addition, I examined how the distribution of messages and firm-days in my sample aligned with the Fama-French 12-industry classifications compared to the broader CRSP universe during the same timeframe. These findings, detailed in Table B.2, reveal that my dataset encompasses all 12 industries, with a pronounced emphasis on technology and healthcare sectors and a lesser representation of financial firms.

Table B.1: Distribution of Posts by Calendar Year

	(1) Firm-Day Observations	(2) Posts
2010	1,337	39,768
2011	3,543	115,710
2012	4,857	295,552
2013	6,718	399,211
2014	11,285	692,239
2015	17,256	893,803
2016	22,734	1,270,722
2017	41,840	2,796,966
2018	53,472	3,453,727
2019	58,645	3,351,474
2020	116,338	9,772,462
2021	141,438	14,575,580
Total	479,463	37,657,214

Table B.2: Distribution of Posts Based on Fama-French 12-Industry Classification

Fama-French industry code (12 industries)	(1) CRSP (%)	(2) Posts (%)	(3) Firm-Day (%)
Consumer NonDurables – Food, Tobacco, Textiles, Apparel, Leather, Toys	4.01	2.73	2.99
Consumer Durables – Cars, TV's, Furniture, Household Appliances	2.30	7.60	3.35
Manufacturing – Machinery, Trucks, Planes, Off Furn, Paper, Com Printing	8.58	3.01	5.04
Oil, Gas, and Coal Extraction and Products	3.13	1.56	2.91
Chemicals and Allied Products	2.38	1.34	1.61
Business Equipment – Computers, Software, and Electronic Equipment	15.01	22.62	21.82
Telephone and Television Transmission	1.75	1.11	1.75
Utilities	2.50	0.31	0.86
Wholesale, Retail, and Some Services (Laundries, Repair Shops)	7.48	5.92	7.58
Healthcare, Medical Equipment, and Drugs	15.13	31.96	33.21
Finance	24.01	4.21	8.23
Other – Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment	13.72	17.63	10.66
Observations	10,357,038	36,518,413	471,307

Notes: CRSP sample follows the same basic sample restrictions; active, traded on U.S. exchanges.

C Examples of StockTwits Model Outputs

Table C.1: Examples of StockTwits Model Outputs

Text	Neutral	Happy	Sad	Anger	Disgust	Surprise	Fear
Financial markets have been uneventful.	79.2%	1.9%	5.3%	1.7%	7.4%	1.3%	3.1%
Today has been such a nice day :).	0.5%	99.0%	0.1%	0.0%	0.0%	0.2%	0.1%
Been a long time since i felt so awful :(.	0.4%	4.1%	81.0%	1.1%	4.5%	3.6%	5.4%
You freaking idiots! Stop selling!!	0.6%	1.2%	3.5%	63.8%	26.0%	2.4%	2.4%
These nasty politicians gotta go!	2.0%	7.0%	1.6%	12.6%	63.8%	0.9%	12.0%
WTf is going on rn??	2.7%	0.3%	0.7%	0.8%	0.6%	93.7%	1.2%
Pretty choppy lately!	22.5%	11.9%	23.0%	3.9%	6.4%	1.9%	30.4%

Notes: Examples of inputs (text) and outputs (i.e., neutral, happy, sad, anger, disgust, surprise, fear) for my StockTwits-based emotion classification model.