CFR Working Paper NO. 25-02

Twitter-Based Attention and the Cross-Section of Cryptocurrency Returns

A. T. Maître • N. Pugachyov • F. Weigert

centre for financial research
cologne

Twitter-Based Attention and the Cross-Section of Cryptocurrency Returns[★]

Arnaud T. Maître[a,*], Nikolay Pugachyov [a] and Florian Weigert [a,1]

[a] *University of Neuchâtel, Institute of Financial Analysis, Abram-Louis-Breguet 2, Neuchâtel 2000, Switzerland*

ARTICLE INFO

Keywords:
Bitcoin
cryptocurrencies
Twitter attention
textual sentiment

ABSTRACT

This paper investigates how investors' abnormal attention affects the cross-section of cryptocurrency returns in the period from 2018 to 2022. We capture abnormal attention using the (log) number of Twitter posts on individual cryptocurrencies on the current day minus a 30-day average. Our results reveal that abnormal attention is positively associated with contemporaneous and one-day ahead crypto performance. Among the different Twitter tweets, return predictability arises due to *Ticker*-tweets from investors, but not due to tweets from the cryptocurrency channel. These *Official*-tweets, however, are able to forecast technological innovations on the blockchain.

1. Introduction

Social media has significant effects on financial markets, as illustrated by the involvement of the WallStreetBets subreddit in the Gamestop short squeeze and the infamous "to the moon" tweets, leading to subsequent Dogecoin price hikes.[1] Internet platforms are particularly important for cryptocurrencies, as they constitute the preferred medium of information exchange between cryptocurrency market participants, with Twitter arguably being the most important platform.[2] To put the importance of social media, and especially Twitter, for cryptocurrencies in context, about 90% of cryptocurrencies have Twitter accounts, whereas this proportion is only 50% for US public firms (Hosseini, Jostova, Philipov and Savickas, 2020).

Given the size of the cryptocurrency market and the significance of Twitter as an information source, studying their relationship is essential for understanding the cross-section of expected cryptocurrency returns.[3] Recent literature provides mixed evidence on the link between Twitter activity and cryptocurrency prices. In the cross-section, Benedetti and Kostovetsky (2021) provide evidence supportive of an overreaction channel consistent with Barber and Odean (2008) and Da, Engelberg and Gao (2011), while Borri, Massacci, Rubin and Ruzzi (2022) identify a negative risk premium for investor attention. In the time-series, Liu and Tsyvinski (2021) and Borri et al. (2022) report a positive relationship between the number of tweets and future cumulative cryptocurrency returns. Altogether, the link between Twitter-based investor attention and cryptocurrency expected returns remains unclear.

Inspired by Da et al. (2011), we inspect the relation between investor attention and cryptocurrency returns and capture abnormal investor attention computed as the (log) number of tweets during the current day minus the (log) mean number of tweets during the previous 30 days. The novelty of our paper is that we collect several different samples of tweets for each cryptocurrency. We consider both tweets written by the organization developing the cryptocurrency (*Official*-tweets) and tweets written by all other users, i.e., user-generated tweets. We further separate user-generated tweets into tweets on the cryptocurrency's ticker (*Ticker*-tweets) and tweets sent to the cryptocurrency's official account (*Mention*-tweets).

Our motivation to categorize tweets is that different categories have unique features, characteristics, and usage motivations. For instance, Barber, Huang, Odean and Schwarz (2022) find that unique features of the Robinhood app partly drive attention-induced trading among the platform's investors. Likewise, *Ticker*-tweets possess the unique feature of being clickable and are commonly employed for discussing trading strategies. Once a user clicks on these tweets, they are presented with the latest tweets related to the associated financial security, making *Ticker*-tweets easily seen by users who are not following the tweet's author. Therefore, *Ticker*-tweets reach a wider audience than the other types of tweets. *Mention*-tweets can be viewed as public messages, as the account tagged by this type of tweet receives a notification. *Official*-tweets, on the other hand, represent corporate announcements. It is particularly interesting to investigate the

[★] We thank Twitter for having provided access to their academic API. We also thank David Ardia and Keven Bluteau for sharing pump and dump data. We thank seminar participants at the 2[nd] International Fintech Conference at Cardiff Business School, the 7[th] Shanghai-Edinburgh-UCL Fintech Conference, the Swiss Society for Financial Market Research (SGF) 2024, the 4[th] Frontiers of Factor Investing 2024 Conference, and University of Neuchâtel for helpful discussions and valuable comments. Any errors are our own.

[*] Corresponding author

arnaud.maitre@unine.ch (A.T. Maître); nikolay.pugachyov@unine.ch (N.P.); florian.weigert@unine.ch (F.W..)

ORCID(s):

[1] Florian Weigert is also affiliated with the Centre of Financial Research (CFR) Cologne and is thankful for their continuous support.

[1] The topic has received wide attention in the empirical finance literature. Previous research shows that social media content has predictive power over expected stock returns and expected earnings (Chen, De, Hu and Hwang, 2014; Bartov, Faurel and Mohanram, 2018; Broadstock and Zhang, 2019; Gu and Kurov, 2020).

[2] The first Bitcoin transaction was arranged on a forum (source) followed by the first-ever Bitcoin-related tweet as early as January 11[th], 2009 (source).

[3] According to Forbes, August 2024, the global cryptocurrency market cap is US$ 2.25 trillion.

possible impact of these features on crypto investors' trading behavior through the lens of attention-induced trading.

In our empirical analysis, we first consider the aggregated set of tweets, i.e., *All*-tweets. For each cryptocurrency, an abnormal attention measure similar to Da et al. (2011) is created. We show that abnormal attention is positively related to contemporaneous returns and next-day returns. On average, a one cross-sectional standard deviation rise in Twitter-based abnormal attention is associated with an increase in cryptocurrency excess returns by 0.70% contemporaneously and 0.11% on the following day. By utilizing our Twitter samples separately, we show that this effect is driven entirely by the user-generated tweets, with *Ticker*-tweets showing a stronger impact than *Mention*-tweets. Thus, unlike Benedetti and Kostovetsky (2021), we do not find a link between *Official*-tweets and excess returns in the cross-section, suggesting that the predictability of Twitter attention arises mainly from user-generated content.

Although the existing literature offers several possible explanations, we show that tweets written by Twitter users predict returns consistently in line with the continued overreaction channel. Boosts in tweet posting activity grab the attention of retail investors who, due to limited cognitive ability, become more likely to purchase these cryptocurrencies (Barber and Odean, 2008; Da et al., 2011). This mechanism results in a temporary positive price pressure. Usually, price rises due to overreaction tend to correct over the following days after the surge in attention, which is not the case in our empirical data. In this regard, Schmeling, Schrimpf and Todorov (2023) point out that limits-to-arbitrage are more important in the crypto market than for other asset classes. Hence, we hypothesize that, e.g., the difficulty of short-selling cryptocurrencies is likely to explain the absence of reversal in returns.

To better understand this difference in predictability power, we create a new lexicon to capture the textual characteristics of user-generated content following the methodology of Renault (2017). We apply this new lexicon to compute a score on the cryptocurrency-day level and show that this score predicts the cross-section of expected returns in the same fashion as abnormal attention. We interpret this finding as evidence that the predictability of user-generated content arises in part from its unique textual content. Furthermore, we show that our results are not driven by the GameStop short-squeeze in 2021, which was surrounded by high social media activity and positive performance for cryptocurrencies.

Given that social media is frequently utilized to report bugs, hacks, or technical problems with blockchain technology, a concurrent explanation of our results is that Twitter abnormal attention predicts returns through its link with future development activities. This hypothesis is motivated by the results of Cong, Li and Wang (2021); Liu, Sheng and Wang (2022a) who reveal that cryptocurrency valuations and ICO success are linked to the quality of their underlying technology. By using the daily number of commits on GitHub as a proxy for technological innovation, we show that *Official*-tweets is the only Twitter sample predicting future technological improvements in the cross-section. As *Official*-tweets do not have predictive power for future returns, the technological innovation channel is not supported by our findings. This difference in predictability between the different Twitter samples supports the notion that social media content is indeed heterogeneous. It also highlights the need to carefully select the appropriate social media data that is best suited for the desired application.

Our paper contributes to two strands of literature. First and foremost, we add to quantify the impact of influential social media users on their followers (Benetton, Mullins, Niessner and Toczynski, 2024; Pedersen, 2022). We document that the posting activity of influential users increases the magnitude of the continued overreaction effect. Our results reveal that even small influencers can have a significant impact on cryptocurrency prices (Benetton et al., 2024). We argue that, similar to media coverage (Hillert, Jacobs and Müller, 2014), influential tweets exacerbate behavioral biases. Therefore, our paper contributes to the nascent literature that explores behavioral biases induced by social media activity.[4]

Second, we add to the emerging literature that studies the impact of information salience on investor behavior (Bose, Cordes, Nolte, Schneider and Camerer, 2022; Kumar, Ruenzi and Ungeheuer, 2021; Frydman and Wang, 2020). For instance, Barber et al. (2022) find that Robinhood users are more likely to purchase stocks displayed in the Top-Movers list than stocks with similar returns absent from the list. In our context, we document that *Ticker*-tweets are seen on average by a wider audience than the other types of tweets after controlling for their number of likes and retweets. Despite this increased visibility of *Ticker*-tweets over *Mention*-tweets, both Twitter samples predict returns similarly in the cross-section. This suggests that our results are not driven by information salience alone. Instead, we provide evidence that the difference in predictability between user-generated content and tweets written by cryptocurrencies is also driven by content differences.

Sections 2 and 3 describe the data and methods used in the paper. We investigate the drivers of Twitter-based attention in Section 4. Section 5 discusses the interplay between aggregate attention and cryptocurrency returns by breaking down attention into several components using the specificity of Twitter. The robustness of our main findings is addressed in Section 6. Finally, Section 7 concludes the paper.

2. Data and Sample Construction

We apply several datasets to construct our sample. Our primary datasets are data on cryptocurrency returns from CoinMarketCap (henceforth CMC) and data on cryptocurrency-related posts on Twitter. We explain the construction of our Twitter sample in detail in Section 2.2. Additionally, we gather data on each cryptocurrency from GitHub, the leading

[4] For example, social media users tend to self-expose to information in line with their beliefs (Cookson, Engelberg and Mullins, 2023) and are influenced by investment returns experienced by users they are following (Bailey, Cao, Kuchler and Stroebel, 2018; Pedersen, 2022).

platform for software development and project collaboration. We use data sampled at a daily frequency in our empirical tests, as Twitter effects documented by the literature are generally short-lived (Benedetti and Kostovetsky, 2021; Gu and Kurov, 2020).

2.1. Cryptocurrency data

CoinMarketCap (henceforth CMC) is a widely accepted data source for cryptocurrency market data. However, the data from the CMC website is subject to survivorship bias because it only provides information on currently listed cryptocurrencies. To address this, we use an API to download survivorship bias-free data from CMC, following the methodology proposed by Ammann, Burdorf, Liebi and Stöckl (2022) that corrects the bias.[5]

Our dataset is at a daily frequency for the period from 2018 to 2022. We drop assets with missing volume data and exclude cryptocurrencies with erroneously reported data. Additionally, we drop stablecoins, which are cryptocurrencies whose value is pegged to other assets such as USD or gold. Our final dataset includes both cryptocoins and tokens.[6] Figure 1 displays the average monthly number of cryptocurrencies meeting our criteria.

2.2. Twitter data

Given the download restrictions imposed by Twitter, we limit our sample to the 165 largest cryptocurrencies as of the end of 2017.[7] We choose this year because it includes a large number of cryptocurrencies meeting our criteria. Our sample excludes cryptocurrencies that have changed names during our sample period or those that do not have a Twitter account.[8] For each selected cryptocurrency, we separately collect tweets written by the organization developing the cryptocurrency (henceforth *Official*-tweets) and tweets written by all other users, i.e., user-generated tweets. While collecting *Official*-tweets is straightforward (simply by requesting tweets posted by the developers' verified accounts), collecting user-generated tweets is more complex. The complexity arises from the fact that users have several ways to signal that their tweet is about a specific cryptocurrency. For instance, users can utilize the name (e.g., #*bitcoin* or #*BTC*), the ticker (e.g., $*BTC*), or tag (mention) the official account of the cryptocurrency (e.g., @*Bitcoin*). We choose to collect tweets on the cryptocurrency's ticker (henceforth *Ticker*-tweets) and tweets sent to the cryptocurrency's official account (henceforth *Mention*-tweets) to exclude tweets with ambiguous hashtags.[9] Ultimately, we categorize tweets into three categories in our Twitter sample: *Official*-tweets, *Mention*-tweets, and *Ticker*-tweets.

Table 1
Description of Twitter samples.
This table presents the different samples of tweets used in the paper along with their number of constituents. The column Description describes which types of tweets are contained in each sample. Note that the Twitter samples have some overlap and are not mutually exclusive, since users can tweet simultaneously about several cryptocurrencies and can mix the types of tweets.

Sample	Description	Number of Cryptos	Defunct Cryptos
Ticker	Any tweet that contains the ticker of the cryptocurrency.	165	44
Mention	Any tweet that tags (mentions) the official account of the crypto.	165	44
Official	Tweets posted by the official account.	154	35
All	The three samples above aggregated.	165	44

Our motivation to categorize tweets is that different categories contain unique features, characteristics, and usage motivations. For instance, Barber et al. (2022) find that unique features of the Robinhood app partly drive attention-induced trading among the platform's investors. *Ticker*-tweets, commonly employed for discussing trading strategies, possess the unique feature of being clickable. When a user clicks on these tweets, they can view the latest tweets related to the associated financial security. Therefore, they have a potentially higher reach than other types of tweets, as they can be easily seen by users who are not following the tweet's author, thanks to the clickable feature. It is particularly interesting to investigate the possible impact of this feature on crypto investors' trading behavior through the lens of attention-induced trading. *Mention*-tweets can be viewed as public messages, as the account tagged by this type of tweet receives a notification. *Official*-tweets, on the other hand, represent corporate announcements.

Table 1 summarizes the information about the sample size and the characteristics of each sample. The number of cryptocurrencies in the *Official* sample is lower than in the other samples because we could not retrieve any *Official*-tweets for eleven cryptocurrencies. We chose to leave the history of *Official*-tweets as missing for those eleven cryptocurrencies because it is hard to tell whether those eleven cryptocurrencies had never posted any tweets during the sample period or had retroactively deleted their tweets.

[5]We access the API through the R package named ***crypto2***, which can be found at https://CRAN.R-project.org/package=crypto2. The authors of this package are Sebastian Stöckl and Jesse Vent. More information about the package can be found on the author's personal website: https://www.sebastianstoeckl.com/

[6]Crypto coins primarily act as a medium of exchange and a store of value. Some well-known examples include Bitcoin (BTC), Ethereum (ETH), and Dogecoin (DOGE). Crypto tokens, on the other hand, are created to serve various purposes, such as utility, ownership, and governance rights. For instance, the Edgeless token can be used to play in an online casino.

[7]Due to the abrupt stop of the Twitter academic API in 2023, we restrict the Twitter samples to the 165 cryptocurrencies for which data could be fully collected.

[8]While it is true that excluding cryptocurrencies that changed their names during the sample period introduces a look-ahead bias, we believe that this bias is unlikely to have a substantial impact on our study's findings. The difference in mean returns between the two samples is not statistically significant from zero.

[9]For instance, consider two cryptocurrencies named ICON (ICX) and TRON (TRX). Both #*icon* and #*tron* can be associated with different meanings, not just cryptocurrencies.

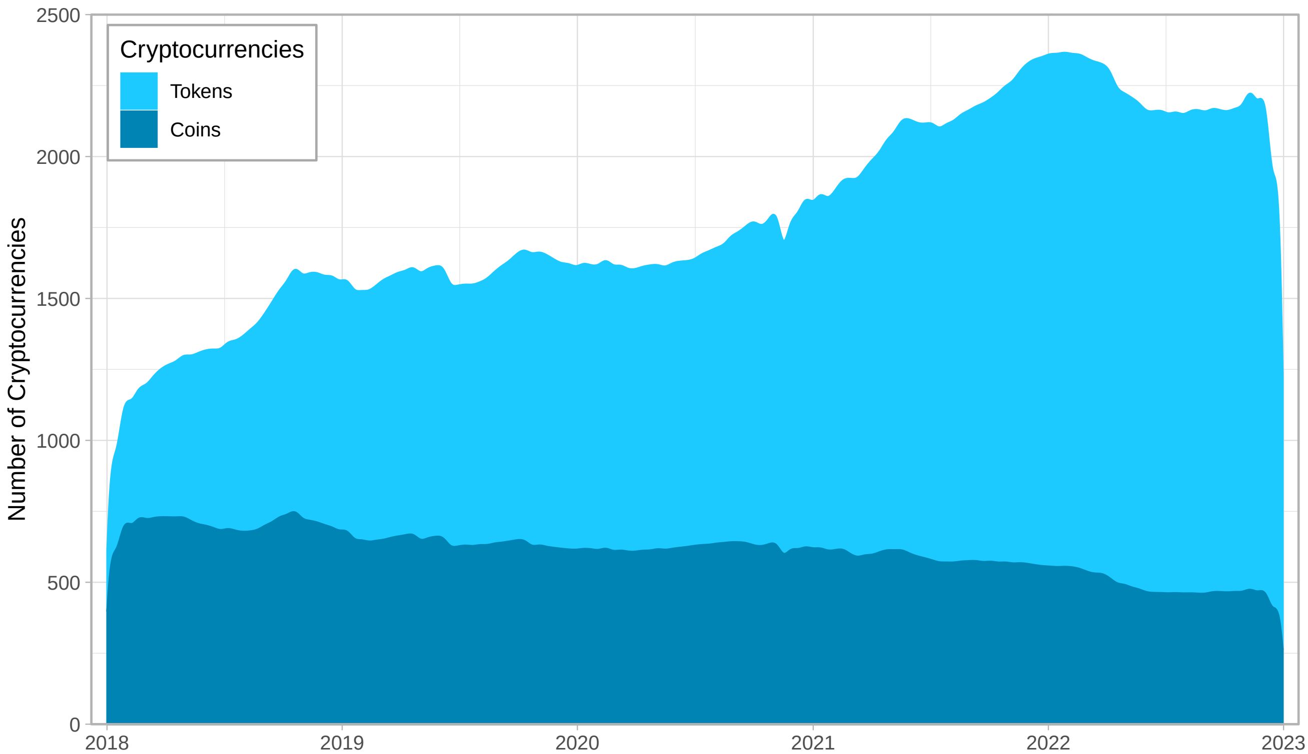

Figure 1: Number of cryptocurrencies, 2018-2022

This figure shows the evolution of the number of cryptocurrencies, coins, and tokens over time at a monthly frequency that meet our criteria. The increase in the number of cryptocurrencies is mainly driven by an increase in tokens. The number of coins has even slightly decreased over the sample period.

Figure 2 reports the number of tweets posted per sample per day. In all our samples, the number of Twitter posts published per day and per cryptocurrency corresponds to the actual number of tweets published on that date. Twitter activity has increased over our sample period, with the only exception being the number of *Official*-tweets, which has remained stable over time. For all samples except the *Official* sample, we only download the 100 most relevant tweets per day over the period from 2018 to 2022.[10] The *Official* sample includes every tweet posted by the cryptocurrency's official account during the sample period.

Unfortunately, as Twitter provides us with tweets based on textual matching of keywords, *Ticker*-tweets may contain measurement error, as tickers are generally not unique. We do not expect this bias to be substantial for two reasons. First, we only consider tweets written in English, which limits the number of ticker homonyms in our sample. Second, as cryptocurrencies are a popular topic on social media, we expect the *Ticker* sample to be primarily composed of tweets about cryptocurrencies.[11] Our Twitter samples are also subject to survivorship bias, as tweets and Twitter accounts can be deleted by their creators. This problem is more severe for the *Official* sample, as it is impossible to retrieve tweets from deleted accounts. Consequently, we miss *Official* tweets from cryptocurrencies with deleted Twitter accounts, which are likely to be defunct. However, also in this case, the bias is assumed to be small, as we could retrieve *Official* tweets from 35 out of the 44 defunct cryptocurrencies included in our Twitter samples.

2.3. GitHub data

To better identify the channels through which Twitter attention predicts the cross-section of cryptocurrency returns, we also collect data on each cryptocurrency from GitHub, the leading platform for software development collaboration. Specifically, we collect the list of all historical contributions (commits) made by developers on all repositories owned by the organization developing the cryptocurrency. Similar to *Official*-tweets, GitHub data is also subject to survivorship bias, as cryptocurrencies with missing data are likely defunct. Our GitHub data covers 143 cryptocurrencies and includes data for 32 out of the 44 defunct cryptocurrencies in our sample.[12]

[10] As determined by Twitter, the exact methodology is not disclosed but considers, among other factors, the degree of keyword matching, tweet engagement, and the author's popularity.

[11] We also report the results if ambiguous tickers are removed from our sample as a robustness check. We judge a ticker as ambiguous if a stock present in the CRSP database has a similar ticker during our sample period. For example, consider a cryptocurrency named Neo (ticker: $*NEO*; https://x.com/neo_blockchain) and a publicly traded company on NASDAQ, NeoGenomics, Inc. (ticker: $*NEO*; https://x.com/NeoGenomics).

[12] We miss data on 22 cryptocurrencies. Three cryptocurrencies use other platforms than GitHub to share their code. We could not find information about 18 cryptocurrencies, and one cryptocurrency's repository contains no commits.

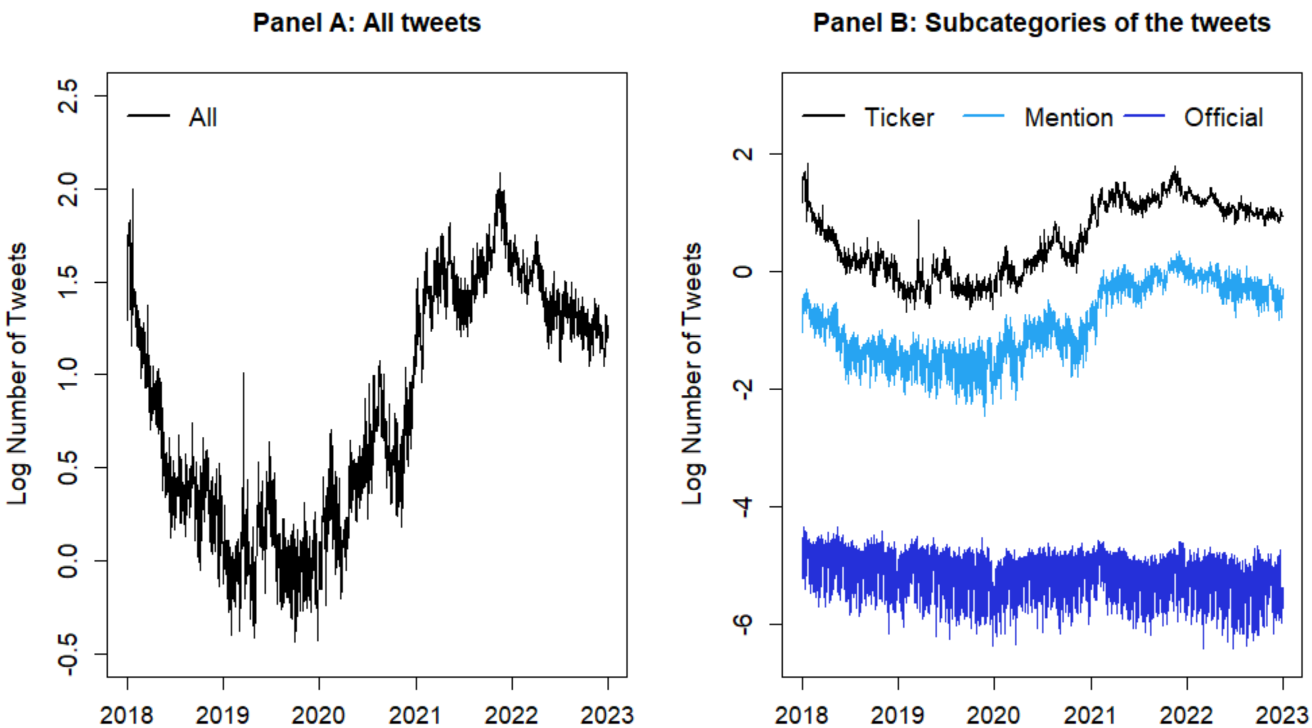

Figure 2: Average number of tweets, 2018-2022

This figure displays the evolution of the average number of tweets for cryptocurrencies over time. The average number of tweets is scaled by the number of cryptocurrencies in our sample at each point in time. For scaling purposes, the series is shown on a logarithmic scale. Panel A displays the evolution of the total number of tweets. Panel B shows the number of tweets separated across the different subcategories.

3. Empirical Methodology

3.1. Abnormal Twitter-based attention

Inspired by Da et al. (2011), we define our main measure of investor attention as abnormal attention, which we compute as the (log) number of tweets during the current day minus the (log) mean number of tweets during the previous 30 days.[13] Specifically,

$$Abn\ Attention_{i,t} = Log(NT_t) - Log(E[NT_{t-30}, ..., NT_{t-1}]), \tag{1}$$

where NT_t is the number of tweets on day t. Abnormal attention has the advantage of being less sensitive to large spikes in the number of tweets than the (log) number of tweets because it takes into account the average level of attention through its rolling mean component.

3.2. Twitter lexicon

The literature has not reached a consensus on which type of social media content is more relevant to empirical studies. Some authors focus on *Ticker*-tweets or other social media platforms oriented towards investing to capture the attention and sentiment of financially savvy users (Chen et al., 2014; Renault, 2017; Ardia and Bluteau, 2024; Gu and Kurov, 2020), whereas others look at content published by companies or aggregated content (Benedetti and Kostovetsky, 2021; Da et al., 2011; Borri et al., 2022). Despite most studies focusing on one type of social media data, there is limited evidence documenting the effects of this decision. To fill this gap, we construct a new lexicon aimed at capturing the differences in textual content between *Ticker*-tweets and other tweets.

We construct our lexicon using a methodology similar to Renault (2017). First, we clean and process the text following common practices. We remove stop words from the tweets and replace words with bigrams when possible, using our own list of bigrams supplemented by the list of Renault (2017).[14] We remove all punctuation. Email addresses, web links, and emojis are replaced by keywords (mailtag, linktag, emojipos, or emojineg).[15] Following Renault (2017), we also add a prefix (negtag) to tokens directly following a negation. Finally, we create a training sample composed of 180,000 tweets equally distributed across *Ticker*-tweets, *Mention*-tweets, and *Official*-tweets.

Importantly, the training sample does not contain any retweets. To avoid any look-ahead bias, the training sample is

[13] Da et al. (2011) rely on Google's Search Volume Index (SVI) to measure investor attention. Their main variable is abnormal SVI, which is defined as the (log) SVI during the current week minus the (log) median SVI during the previous eight weeks.

[14] For forming the bigrams, we use 300,000 randomly selected tweets over the complete sample.

[15] We manually classify the most frequent emojis as either being positive or negative.

only composed of tweets published in 2017. We also filter out irrelevant unigrams and bigrams by restricting our sample to n-grams appearing in at least 0.01% of tweets. To prevent our final lexicon from being too influenced by prolific authors, cryptocurrencies, or seasonal topics, we require n-grams to be used by at least 10 different authors, to be used in discussions about at least 10 different cryptocurrencies, and to be used in at least three distinct months. Our last filtering step is to remove tickers and user mentions from the selected tokens to prevent our lexicon from loading on terms that are specific to one sample of tweets.

To spot unigrams and bigrams that are specific to *Ticker*-tweets, we compute a coefficient for each token i which is defined as:

$$C_i = \frac{Occurence_i^T - \max(Occurence_i^M; Occurence_i^O)}{Occurence_i^T + \max(Occurence_i^M; Occurence_i^O)}, \quad (2)$$

where T, M, and O stand for Ticker, Mention, and Official, respectively.

Like Renault (2017), we only keep in the lexicon n-grams with a coefficient in the first or the last quintile. The resulting lexicon is composed of 2,833 n-grams. Table 2 presents a selection of the lexicon's vocabulary. We observe that the lexicon loads positively on terms related to trading such as 'bull–flag' or 'sold–all' regardless of their sentiment and negatively on terms related to technology or news such as 'support–team' or 'fully-decentralized'.

The lexicon confirms the intuition that different types of users use different types of tweets. Users tend to use mainly *Ticker*-tweets to discuss cryptocurrency investing. In contrast, *Official*-tweets and *Mention*-tweets are used for corporate announcements and to signal and solve issues related to the cryptocurrency. In the rest of the paper, we compute the lexicon score per tweet as the equal-weighted average coefficient of all n-grams present in the tweet that belong to the lexicon. We then compute the cryptocurrency-day lexicon score as the average score of all tweets published about that cryptocurrency on a given date. To illustrate our methodology, the hypothetical tweet '*@username please contact support team*' would get a score of $-0.8189 = \frac{-0.7353+-0.9024}{2}$ where -0.7353 and -0.9024 are respectively the score assigned by the lexicon for the n-grams 'contact' and 'support-team'.

3.3. Variables description and summary statistics

As in our Twitter data, we observe the number of likes, text, and author identity for a subset of the tweets posted per cryptocurrency-date; however, we are not able to identify the complete activity of tweet authors. Consequently, we choose to measure influential user activity in an indirect manner by checking if at least one influential user has tweeted on this cryptocurrency-date. We construct a dummy variable ($PopularTweet_{i,t}^{All}$) equal to one when at least one tweet published per cryptocurrency-date has an aggregated number of likes and replies equal to or above 100 across all of our Twitter samples.[16] The number 100 corresponds to the 93[th] percentile of the distribution of the sum of the number of comments and likes. Classifying tweets with as few as 100 likes and/or comments as influential might seem optimistic. However, the actual number of users reached by a tweet is often much larger than its number of likes and replies. Using the number of views of each tweet, we find that the median number of views for tweets with at least 100 likes is 17,523 views and this number does not include the views obtained by the activity induced by popular tweets.[17] We choose to compute this variable across all Twitter samples simultaneously because we are interested in how influential users' activity, regardless of their tweeting preferences, impacts other Twitter users.

Sentiment- and lexicon-based variables are constructed at the tweet level and then averaged at the cryptocurrency-date level. To compute sentiment, we use a bag-of-words approach and consider the dictionary of Renault (2017), which is the main lexicon used for quantifying the sentiment of texts from social media in the financial academic literature.[18]

We follow Liu, Tsyvinski and Wu (2022b) to construct the cryptocurrency market factor. Throughout the paper, we

Table 2
Selected n-grams from lexicon.
This table contains selected n-grams from the lexicon along with their number of occurrences across the different Twitter samples.

N-gram count:	*Mention*	*Official*	*Ticker*	Coeff.
answering–questions	2	16	0	-1.00
full–stack	6	12	0	-1.00
sorry–inconvenience	5	35	1	-0.94
patch	9	32	1	-0.94
thank–patience	4	30	1	-0.93
assistance	16	99	4	-0.92
support–team	7	39	2	-0.90
pull–request	8	33	2	-0.89
token–sale	53	371	29	-0.86
opensource	49	122	13	-0.81
pleased–announce	22	42	5	-0.79
source–code	13	24	6	-0.60
vulnerabilities	6	11	3	-0.57
fully–decentralized	7	11	3	-0.57
withdrawing	4	11	3	-0.57
trading	522	669	1626	0.42
sold–all	11	0	28	0.44
take–profits	7	0	26	0.58
sell–orders	4	6	24	0.60
still–cheap	14	0	59	0.62
buy–hold	15	2	66	0.63
resistance	46	29	394	0.79
daily–chart	2	0	21	0.83
overbought	2	0	24	0.85
pattern	11	6	166	0.88
buys–numbertag	2	2	36	0.89
stop–loss	4	0	83	0.91
bull–flag	1	0	46	0.96
made–returntag	10	3	692	0.97
break–above	0	0	25	1.00

[16]The number of retweets is excluded from this total. We do so because the number of retweets is shared across the retweets and the original tweet. Therefore, using the number of retweets to spot popular tweets can be misleading.

[17]We do not directly use the number of views to spot popular tweets in our analysis because this variable is only available for a limited number of tweets.

[18]Our results do not change if we use the lexicon of Loughran and McDonald (2011) instead.

control for a wide range of variables impacting asset prices. We control for illiquidity and liquidity fluctuations using the measure of Amihud (2002); Chordia, Subrahmanyam and Anshuman (2001) and measures of codependence with the cryptocurrency market using co-skewness (Kraus and Litzenberger, 1976; Harvey and Siddique, 2000) and co-kurtosis (Fang and Lai, 1997; Jondeau and Rockinger, 2006). In addition, we also include controls related to volatility, higher-order moments of returns, and tail risk using variables such as idiosyncratic volatility (Ang, Hodrick, Xing and Zhang, 2006), skewness, kurtosis, and value-at-risk.

Momentum is included in our set of controls, as momentum strategies are profitable for a wide range of asset classes (Asness, Moskowitz and Pedersen, 2013) and are related to investor attention (Daniel, Hirshleifer and Subrahmanyam, 1998; Hillert et al., 2014). Hillert et al. (2014) find that media coverage of larger firms attracts investor attention (Solomon, Soltes and Sosyura, 2014) and intensifies investors' overconfidence and self-attribution biases (Daniel et al., 1998), which result in temporary improvements in momentum returns. However, evidence from Liu and Tsyvinski (2021) suggests that the relationship between momentum returns and investor attention is different between cryptocurrencies and stocks. The authors find that the return predictability of attention and momentum does not encompass each other.

In addition, we add a control variable capturing the 'MAX effect' (Bali, Cakici and Whitelaw, 2011; Fong and Toh, 2014) which is related to investor sentiment and gambling preferences (Fong and Toh, 2014). The GameStop short-squeeze attracted a significant number of retail speculators and triggered a surge in investor attention (Lyócsa, Baumöhl and Vỳrost, 2022). This event suggests that the 'Max effect' and investor attention could be related. Table 3 provides a description of all variables used in our paper.

Given the suspicious returns documented by Ammann et al. (2022), we follow their approach and trim the daily cryptocurrency returns at the 99% level. The remaining variables are winsorized at the 1% level. Table 4 displays descriptive statistics for the variables used in our paper. In our data, abnormal attention is negative on average and is negatively skewed across all samples. Furthermore, we can see that tweets tend to have positive sentiment and to focus on terms related to trading as indicated by the positive mean of sentiment and lexicon scores. In terms of popularity on Twitter, cryptocurrencies differ significantly. For instance, Bitcoin gets a median number of tweets of 11,236 per day, while the median across all cryptocurrencies is 60 tweets. The low number of tweets for some assets is not purely driven by size; the smallest cryptocurrency we consider as a market cap of $47 million as of the start of our sample. Even if the median number of tweets seems low, one has to remember that the actual number of people reached by tweets is larger. The median number of views per tweet is about 118 and it does not include the number of times that tweets were collected by web-scrapping algorithms. Anecdotal evidence suggests that automatized data collection efforts can be sizeable. For instance, 27% of hedge funds surveyed by Ernst and Young in their 'EY Global Hedge Fund and Investor Survey 2017' reports that they are using or planning to use social media data as part of their investment process.[19] Furthermore, the low popularity of some cryptocurrencies goes against us finding any return predictability for Twitter attention.

4. Characterizing Twitter-Based Investor Attention

4.1. Drivers of Twitter-based investor attention

We start our empirical analysis by investigating the drivers of Twitter-based investor attention. For this purpose, Table 5 reports the results on contemporaneous relationships between the different attention measures. We use changes in attention rather than abnormal attention because we are interested in what causes variation in attention more generally. The panel regression models control for the variables defined in Panel A of Table 3.

Table 5 reveals that attention variables are positively related to each other at the 1% or 5% significance level. Daily excess returns are also significantly linked to contemporaneous changes in attention for *Mention*-tweets and *Ticker*-tweets at the 1% level, but do not appear to be linked to the change in the number of *Official*-tweets. The presence of $PopularTweet_{i,t}^{All}$ is strongly positively related to changes in the number of tweets across all three samples. This is in line with the results of Benetton et al. (2024) who document that celebrity tweeting activity influences their followers' behavior. Variables derived from volume, $\Delta Volume_{i,t}$ and $VolumeVola_{i,t}$, are significantly linked with the dependent variables. Changes in volume have a similar regression coefficient to excess returns, and volatility of volume is negatively linked with changes in the number of tweets from each sample. Interestingly, momentum is negatively related to attention changes, which suggests that the association between attention and momentum is less strong for cryptocurrencies compared to equities (Liu and Tsyvinski, 2021). We do not find strong links between change in tweeting activity and the other control variables described in Panel A of Table 3.

We run similar tests using future change in tweeting activity in Table 6. Future changes in the different changes in attention measures are inversely related to their own one-day lags at the 1% significance level. Interestingly, we observe differences in cross-one-day auto-correlations between attention variables. For instance, *Official*-tweets and *Ticker*-tweets positively predict future changes in the number of tweets for the other sample at the 1% significance level. On the other hand, *Mention*-tweets are negatively related to future changes in *Official*-tweets. As in Table 5, daily cryptocurrency returns and volume changes are strong predictors of the number of tweets, even for *Official*-tweets. Furthermore, our results reveal that coefficient estimates of $PopularTweet_{i,t}^{All}$ become negative. We interpret these results as indicative that attention tends to revert to a normal

[19] https://eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2017/12/ey-how-will-you-embrace-innovation.pdf

Table 3
Variables definition.
This table contains a description of the variables used in the paper. CMC stands for CoinMarketCap's website. "Author Homepage" indicates that the data described in the variable definition can be found on the website of the respective authors. KF is Kenneth French's website.

Variable	Definition	Source
Panel A: Cryptocurrency returns and characteristics		
$Excess\ Return_{i,t}$	Excess return on day t for cryptocurrency i.	CMC, KF
$Size_{i,t}$	Logarithmic market capitalization on day t for cryptocurrency i.	CMC
$\Delta Volume_{i,t}$	Logarithmic daily change in trading volume on day t for cryptocurrency i.	CMC
	The variables listed below until the start of Panel B are all computed over a rolling window of 60 days with a minimum of 30 days of non-missing observations.	
$Beta_{i,t}$	Regression coefficient of daily excess return of cryptocurrency i on the daily cryptocurrency market excess return.	CMC
$Momentum_{i,t}$	Compounded return of cryptocurrency i.	CMC
$Volatility_{i,t}$	Standard deviation of returns of cryptocurrency i.	CMC
$IdioVola_{i,t}$	Standard deviation of the residuals when daily excess returns of cryptocurrency i are regressed on daily cryptocurrency market excess returns.	CMC
$Max\ Ret_{i,t}$	Average of the five highest daily excess return of cryptocurrency i.	CMC
$VolumeVola_{i,t}$	Standard deviation of Log transformed trading volume of cryptocurrency i.	CMC
$Illiquidity_{i,t}$	Ratio of illiquidity of cryptocurrency i, see Amihud (2002). $Amihud = \frac{1}{T}\Sigma_t^T \frac{\lvert r_{i,t}\rvert}{Volume_{i,t}}$	CMC
$Skewness_{i,t}$	Skewness of cryptocurrency i daily excess returns.	CMC
$Kurtosis_{i,t}$	Kurtosis of cryptocurrency i daily excess returns.	CMC
$Co-skewness_{i,t}$	The Co-Skewness of cryptocurrency i daily excess returns with daily cryptocurrency market excess return. $Coskew = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^2]}{\sigma_{r_i}\sigma^2_{r_m}}$	CMC
$Co-kurtosis_{i,t}$	The Co-Kurtosis of cryptocurrency i daily excess returns with daily cryptocurrency market excess return. $Cokurt = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^3]}{\sigma_{r_i}\sigma^3_{r_m}}$	CMC
$VaR_{i,t}$	The 5% percentile of daily cryptocurrency i excess returns.	CMC
Panel B: Attention and sentiment measures		
$Abn\ Attention^{All}_{i,t}$	Difference between the (log) number of *All*-tweets of cryptocurrency i at time t and the (log) mean number of *All*-tweets of cryptocurrency i during the previous 30 days. See equation 1. Abnormal attention for other Twitter samples is defined analogously.	Twitter
$\Delta Attention^{All}_{i,t}$	Difference between the (log) total number of *All*-tweets of cryptocurrency i at time t and the (log) number of *All*-tweets of cryptocurrency i at time $t-1$.	Twitter
$Sentiment^{All}_{i,t}$	Sentiment of the tweets published at date t on cryptocurrency i. We use the Renault (2017) lexicon to compute the sentiment of tweets. Sentiments of individual tweets are then averaged to get a sentiment score at a daily frequency. Sentiment is computed using the sample *All*-tweets.	Twitter, Author Homepage
$PopularTweet^{All}_{i,t}$	Dummy variable equals one if at least one tweet from any sample published about cryptocurrency i at time t has an aggregated number of likes and replies equal to or higher than 100 across all Twitter samples. $PopularTweet^{All}_{i,t}$ is computed using the sample *All*-tweets.	Twitter
$Lexicon^{All}_{i,t}$	Lexicon score of the tweets published at date t on cryptocurrency i. The lexicon is made to capture terms that are specific to *Ticker*-tweets. Lexicon score is computed using the sample *All*-tweets.	Twitter

Table continued on next page

level following large changes; Holding other factors equal, a *Popular*-tweet posted at time t stimulates the number of tweets posted at time t and reinforces the attention reversal at time $t+1$. Similarly to Table 5, momentum is negatively related to future changes in tweeting activity in all models.

Taken together, Table 5 and Table 6 both suggest that tweeting activity tends to spike and revert partially over the following day. As expected, the link between past returns and Twitter attention is sizeable (Da et al., 2011; Liu and Tsyvinski, 2021) and is the strongest for *Ticker*-tweets, both in terms of statistical and economic significance. Intuitively, this observation makes sense given that *Ticker*-tweets are mainly used for discussing trading and investments as shown in Table 2. Despite Twitter users reacting to past and contemporaneous

Table continued

Panel C: Tweet characteristics		
$RT_{i,n,t}$	(Log) sum of the number of retweets and the number of quotes + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$PM_{i,n,t}$	(Log) sum of the number of likes and the number of bookmarks + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$Views_{i,n,t}$	(Log) number of times the tweet has been seen on Twitter + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$LengthTweet_{i,n,t}$	(Log) length of the tweet + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
Panel D: Technology improvement measures		
$\Delta Commit_{i,t}$	(Log) daily change of the number of commits + 1 published on GitHub at time t for each repository of the organization developing the respective cryptocurrency i.	GitHub

returns, change in attention is inversely related to momentum as documented by Liu and Tsyvinski (2021). The authors argue that if momentum arises from underreaction to news (Hong and Stein, 1999), then momentum and change in attention should be negatively correlated as observed in our data.

While the different attention variables are positively related contemporaneously, we observe different patterns at $t+1$, especially for *Mention*-tweets, which become negatively related to the other samples. One could wonder why *Official*-tweets and *Ticker*-tweets are positively associated with future tweeting activity in other samples. We posit that those two samples of tweets are on average more visible and therefore induce more people to tweet. *Ticker*-tweets have a unique clickable feature which should increase their visibility, hence their impact. Alternatively, *Official*-tweets are posted by a recognized Twitter user and constitute an important source of information for cryptocurrency investors. We investigate our hypothesis in the following section.

4.2. Differences between the Twitter samples

The three categories of tweets used in this study have different characteristics. Their most important difference is how they are disseminated on Twitter. By default, a tweet is displayed on the profile page of the author and in the timeline of users following the author.[20] However, both *Ticker*-tweets and *Mention*-tweets differ from this default behavior. *Ticker*-tweets also appear in the respective financial security timeline. *Mention*-tweets can either follow the default behavior or be customized such that these tweets only appear in the timeline of users following both the author and the user being tagged. Therefore, *Ticker*-tweets should be more visible than classic tweets, whereas *Mention*-tweets should be less visible.

Several studies show that information salience impacts investor behavior (Barber and Odean, 2008; Barber et al., 2022). For instance, Tan, Wang and Zhou (2015) find that better readability helps investors to better incorporate new information. Therefore, it is reasonable to expect that the Twitter samples predict returns differently in the cross-section based on their visibility. As a first step, we test in Table 7 whether the category of the tweet indeed impacts its visibility and other tweet metrics. The different models use date and cryptocurrency fixed effects to control for unobserved invariant characteristics and cluster standard errors by authors. In addition, we exclude retweets from each estimated model. As original tweets and their retweets all share the same retweet count, keeping retweets in the sample may lead to spurious relationships.

Our results reveal that being a *Ticker*-tweet or an *Official*-tweet positively relates to the number of views obtained by the tweet. This effect is statistically significant at the 1% and 5% levels for *Official*-tweets and *Ticker*-tweets, respectively. In addition, *Ticker*-tweets are also significantly associated with a larger number of retweets. In contrast, we find a negative relationship between *Mention*-tweets and the number of views obtained by the tweet at the 1% level. Textual sentiment is positively related to the number of likes, replies, and retweets at the 1% significance level, but negatively correlated with the number of views at the 1% level. The lexicon score, which captures trading jargon, is positively linked with the number of views at the 1% significance level, but negatively associated with the other dependent variables at the 1% significance level.

The results support our intuition that *Ticker*-tweets are more visible than classic tweets thanks to their clickable feature. In addition, we find similar results for *Official*-tweets, which is in line with our hypothesis that investors value the news announced by *Official*-tweets. In contrast, the regression coefficients of *Mention*-tweets in the last model are negative. As expected, *Mention*-tweets are generally written with the intent to reach a specific set of users. Interestingly, $Lexicon_{i,n,t}$ is positively associated with the number of views and negatively associated with the number of retweets and likes. Social media users do not seem to actively share tweets with high lexicon scores, but they still search for them as indicated by the positive correlation with the number of views. We observe similar patterns for the effects of tweets with negative sentiment on tweet visibility. One potential interpretation could be that Twitter users actively search for bad news and other users' opinions as part of their investment process. Such behavior would be in line with the 'Do your own research' (DYOR) advice which is frequently addressed to new cryptocurrency investors on social media.[21] Overall, the evidence contained in Table 7 confirms that the different

[20] https://help.twitter.com/en/using-x/types-of-posts

[21] The unigram 'dyor' has a score of 0.5673 in the lexicon.

Table 4
Summary statistics.
This table contains the summary statistics of the variables defined in Table 3. All variables are winsorized at the 1% level, except for $Excess\ Returns_t$ which is trimmed at the 99% level. The variables are expressed in decimal points. All variables are at a daily frequency.

	Mean	25%	Median	75%	StdDev
Panel A: Cryptocurrency returns and characteristics					
$Excess\ Return_{i,t}$	−0.0006	−0.038	−0.002	0.032	0.092
$Beta_{i,t}$	0.93	0.76	0.96	1.14	0.35
$Size_{i,t}$	17.43	15.75	17.12	18.84	2.53
$Momentum_{i,t}$	0.008	−0.422	−0.161	0.175	0.780
$Volatility_{i,t}$	0.080	0.052	0.069	0.096	0.041
$Idio\ Vola_{i,t}$	0.064	0.035	0.051	0.079	0.044
$Max\ Ret_{i,t}$	0.161	0.096	0.134	0.199	0.094
$\Delta Volume_{i,t}$	−0.005	−0.308	−0.023	0.255	0.731
$Volume\ Vola_{i,t}$	0.783	0.462	0.689	0.968	0.481
$Illiquidity_{i,t}$	0.002	0	0	0	0.021
$Skewness_{i,t}$	0.375	−0.216	0.280	0.872	1.025
$Kurtosis_{i,t}$	6.173	3.612	4.723	7.042	4.073
$Co\text{-}Skewness_{i,t}$	−0.345	−0.575	−0.279	−0.034	0.515
$Co\text{-}Kurtosis_{i,t}$	3.108	1.701	2.569	3.667	2.509
$VaR_{i,t}$	−0.113	−0.131	−0.099	−0.076	0.062
Panel B: Attention and sentiment measures					
$Abn\ Attention^{All}_{i,t}$	−0.210	−0.571	−0.179	0.176	0.675
$Abn\ Attention^{Mention}_{i,t}$	−0.315	−0.790	−0.210	0.111	0.834
$Abn\ Attention^{Official}_{i,t}$	−0.119	−0.383	−0.065	0.034	0.487
$Abn\ Attention^{Ticker}_{i,t}$	−0.201	−0.546	−0.162	0.166	0.665
$Sentiment^{All}_{i,t}$	0.041	0	0.044	0.084	0.062
$Popular\ Tweet^{All}_{i,t}$	0.222	0	0	0	0.416
$Lexicon^{All}_{i,t}$	0.097	−0.014	0.099	0.236	0.184
$\#Tweets^{All}_{i,t}$	812	16	60	233	5127
Panel C: Tweet characteristics					
$RT_{i,n,t}$	1.594	0	1.099	2.565	1.657
$PM_{i,n,t}$	1.004	0	0	1.609	1.380
$Views_{i,n,t}$	4.547	2.773	4.771	6.528	2.934
$Length\ Tweet_{i,n,t}$	2.613	2.197	2.639	3.044	0.553
Panel D: Technology improvement measures					
$\Delta Commit_{i,t}$	−0.0008	−0.29	0	0.24	0.90

characteristics of tweets matter and impact their salience on Twitter, thus providing evidence that the choice of social media is not innocuous.

Table 5
Contemporaneous determinants of refined Twitter-based attention.
The dependent variable is the change in attention using several Twitter samples. The regression spans the period from 2018 to 2022 for a sample of 154 cryptocurrencies. Control variables are defined in Panel A of Table 3. The regression coefficients are expressed in percentage points. Standard errors are clustered along time and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1) $\Delta Att^{Mention}_t$	(2) $\Delta Att^{Official}_t$	(3) ΔAtt^{Ticker}_t
$\Delta Attention^{Mention}_{i,t}$		22.95***	10.34***
		(26.33)	(16.75)
$\Delta Attention^{Official}_{i,t}$	46.46***		1.3**
	(21.75)		(2.54)
$\Delta Attention^{Ticker}_{i,t}$	20.84***	1.3**	
	(15.33)	(2.53)	
$Excess\ Return_{i,t}$	24.39***	−2.23	53.75***
	(8.44)	(−1.55)	(10.93)
$Popular\ Tweet^{All}_{i,t}$	24.77***	12.05***	14.38***
	(12.01)	(9.23)	(13.38)
$Beta_{i,t}$	−0.46	0.22	−0.44
	(−1.35)	(0.82)	(−1.07)
$Size_{i,t}$	−1.55***	−0.72***	−0.96***
	(−5.5)	(−4.13)	(−4.87)
$Momentum_{i,t}$	−0.32**	−0.34***	−0.69***
	(−2.12)	(−3.54)	(−3.43)
$Volatility_{i,t}$	−18.81	−24.29	14.55
	(−0.85)	(−1.55)	(0.6)
$Idio\ Vola_{i,t}$	10.07	25.32*	3.93
	(0.59)	(1.88)	(0.2)
$Max\ Ret_{i,t}$	−2.81	1.01	−9.37**
	(−0.77)	(0.51)	(−2.11)
$\Delta Volume_{i,t}$	4***	−0.39*	10.36***
	(9.9)	(−1.95)	(12.03)
$Volume\ Vola_{i,t}$	−0.9***	−0.37***	−0.91***
	(−3.33)	(−2.71)	(−3.93)
$Illiquidity_t$	−7.15	−5.65	−1.02
	(−0.8)	(−1.52)	(−0.14)
$Skewness_{i,t}$	0.21	0	0.24**
	(1.26)	(0.04)	(2.12)
$Kurtosis_{i,t}$	0	−0.01	−0.06**
	(0.15)	(−0.45)	(−2.16)
$Co-Skewness_{i,t}$	−0.79**	−0.21	0.14
	(−2.52)	(−0.79)	(0.45)
$Co-Kurtosis_{i,t}$	0.2*	0.05	0.25**
	(1.84)	(0.6)	(2.36)
$VaR_{i,t}$	−7.16*	2.34	2.56
	(−1.68)	(1.04)	(0.68)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	246957	246957	246957

5. Twitter-Based Investor Attention and the Cross-Section of Cryptocurrency Returns

5.1. Twitter-based abnormal attention

This section studies the link between abnormal attention and the cross-section of cryptocurrency returns. In Table 8, we present the results of panel regressions of excess returns on abnormal attention and various controls as defined in Panel A of Table 3. Model (1) investigates the association between Twitter attention and returns contemporaneously. The other models replace contemporaneous returns with future returns using different time horizons. One potential issue with *Ticker*-tweets is that their trading symbol is not unique, as trading symbols are attributed by the exchange where assets are traded. Given that cryptocurrencies are traded on their own exchanges, they can possibly share their ticker with other financial securities traded elsewhere such as US stocks. Therefore, for robustness, we replicate the same regressions with a different sample in Panel B; when doing so, we drop cryptocurrencies that share their ticker with a firm covered by the CRSP database during our sample period.[22]

[22] The cryptocurrencies that are dropped in Panel B include both large and small cryptocurrencies. For instance, both Bitcoin and Ethereum are dropped in Panel B.

Table 6
Predictive determinants of refined Twitter-based attention.
The dependent variables are changes in attention using several Twitter samples. The regression spans the period from 2018 to 2022 for a sample of 154 cryptocurrencies. Control variables are defined in Panel A of Table 3. The regression coefficients are expressed in percentage points. Standard errors are clustered along time and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1) $\Delta Att_{t+1}^{Mention}$	(2) $\Delta Att_{t+1}^{Official}$	(3) $\Delta Att_{t+1}^{Ticker}$
$\Delta Attention_{i,t}^{Mention}$	−36.6***	−1.58***	−0.08
	(−82.21)	(−9.04)	(−0.47)
$\Delta Attention_{i,t}^{Official}$	10.33***	−41.78***	1.31***
	(15.02)	(−78.76)	(5.4)
$\Delta Attention_{i,t}^{Ticker}$	2.44***	0.66***	−32.52***
	(6.22)	(2.83)	(−66.33)
$Excess\ Return_{i,t}$	19.56***	2.81**	25.93***
	(6.91)	(2.19)	(8.06)
$Popular\,Tweet_{i,t}^{All}$	−7.38***	−9.97***	−2.99***
	(−7.52)	(−11.26)	(−6.58)
$Beta_{i,t}$	−0.35	0.08	−0.66*
	(−0.97)	(0.24)	(−1.76)
$Size_{i,t}$	−0.05	0.51***	−0.36**
	(−0.31)	(2.93)	(−2.27)
$Momentum_{i,t}$	−0.73***	−0.2*	−1.31***
	(−4.13)	(−1.85)	(−5.95)
$Volatility_{i,t}$	44.97**	18.2	33.2
	(2.06)	(0.99)	(1.6)
$Idio\,Vola_{i,t}$	−50.16***	−15.43	−35.68**
	(−2.99)	(−0.93)	(−2.09)
$Max\ Ret_{i,t}$	−2.96	0.73	−2.68
	(−0.64)	(0.34)	(−1.03)
$\Delta Volume_{i,t}$	1.9***	0.48***	1.57***
	(6.95)	(2.7)	(7.24)
$Volume\,Vola_{i,t}$	0.09	0.18	−0.22
	(0.58)	(1.24)	(−0.96)
$Illiquidity_t$	7.73	7.37**	1.84
	(1.51)	(2.01)	(0.32)
$Skewness_{i,t}$	0.11	−0.1	0.08
	(0.63)	(−1.09)	(0.7)
$Kurtosis_{i,t}$	−0.07***	−0.02	−0.09***
	(−2.62)	(−0.99)	(−4.12)
$Co-Skewness_{i,t}$	0.08	0.19	0.61*
	(0.27)	(0.74)	(1.82)
$Co-Kurtosis_{i,t}$	0.1	−0.06	0.22**
	(0.98)	(−0.68)	(2.01)
$VaR_{i,t}$	−1.06	3.07	−0.36
	(−0.24)	(1.36)	(−0.1)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	246844	246844	246844

Table 7
Visibility of tweets.
The dependent variables are the number of retweets, the number of likes and replies, and the number of views for each tweet. The dependent variables are log-transformed. The sample covers the period from December 15, 2022, to December 31, 2022, for a sample of 151 cryptocurrencies. The regression coefficients are expressed in percentage points. Standard errors are clustered by authors. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1) $RT_{i,n,t}$	(2) $PM_{i,n,t}$	(3) $Views_{i,n,t}$
$Mention\,Tweet_{i,n,t}$	0.32	0.04	−34.12***
	(0.08)	(0.01)	(−6.09)
$Official\,Tweet_{i,n,t}$	105.59***	7.45	50.21***
	(10.84)	(1.19)	(6.64)
$Ticker\,Tweet_{i,n,t}$	31.75***	3.5	9.89**
	(8.62)	(1.05)	(2.05)
$RT_{i,n,t}$		41.24***	−65.53***
		(63.18)	(−64.82)
$PM_{i,n,t}$	77.03***		158.85***
	(59.06)		(104.52)
$Views_{i,n,t}$	−29.47***	38.25***	
	(−50.86)	(118.99)	
$Length\,Tweet_{i,n,t}$	21.07***	−4.64***	25.87***
	(15.64)	(−2.7)	(7.66)
$Sentiment_{i,n,t}$	20.41***	16.04***	−79.69***
	(5.5)	(4.71)	(−11.84)
$Lexicon_{i,n,t}$	−13.72***	−10.62***	28.64***
	(−8.08)	(−5.11)	(7.95)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	63221	63221	63221

Our results reveal that the regression coefficient of $Abn\ Attention_t^{All}$ is statistically significant and positive in models (1) and (2) at the 1% significance level. Evaluating its economic significance, the effect is large; a one standard-deviation increase in $Abn\ Attention_t^{All}$ is related to an increase of 0.72% in contemporaneous daily excess returns and an increase in future daily excess returns of 0.11%. The association between abnormal attention at day t and excess returns at $t+1$ hence amounts to an annual 39%. The predictability of Twitter activity is robust to a wide range of popular predictors used in the literature. Among those predictors, only past returns, size, and change in volume are all significantly related to contemporaneous and future cryptocurrency returns. Volume change predicts returns in a similar way as abnormal attention. Past returns predict future returns up to two periods ahead, and size is negatively related to returns in all model specifications, consistent with previous literature. Furthermore, we document similar results in Panel B where cryptocurrencies with ambiguous trading symbols have been removed, suggesting that our results are not driven by tweets that should not be included in our sample.

The positive connection between $Abn\ Attention_t^{All}$ and excess returns is consistent with the results of Liu and Tsyvinski (2021), who also document a positive relationship between investor attention and cryptocurrency returns. However, we need to be cautious when interpreting the results of Table 8, as several theories could explain why attention and (expected) returns are positively correlated. For instance, as Twitter is used to discuss the technology underlying cryptocurrencies, Twitter attention could predict technological improvements and therefore returns (Cong et al., 2021; Lyandres, Palazzo and Rabetti, 2022; Liu et al., 2022a). Another possibility is that attention predicts increases in the user base, which would have positive network externalities for existing cryptocurrency users (Cong et al., 2021; Sockin and Xiong, 2023).

Alternatively, the results could also be explained by an overreaction channel (Barber and Odean, 2008; Da et al., 2011). Increases in tweet posting activity grab the attention of retail investors, who, due to limited cognitive ability, become more likely to purchase those attention-grabbing cryptocurrencies. This mechanism results in temporary positive price pressure. Usually, price rises caused by overreaction tend to correct over the following days after the surge in attention. However, we do not observe such price correction (as revealed in models (3) and (4)). In this regard, Schmeling et al. (2023) point out that limits-to-arbitrage are important in the cryptocurrency market and prevent some arbitrage strategies, which are commonly used for other asset classes, from being profitable for cryptocurrencies. The difficulty of short-selling cryptocurrencies could therefore explain the absence of reversal.[23]

As documented in Tables 2 and 7, each Twitter sample has some unique properties; Tweets are not equally salient between our samples, nor do they share similar textual content. Moreover, the samples also exhibit qualitative differences. For instance, *Official*-tweets are mainly constituted of announcements about their associated cryptocurrency and thus differ fundamentally from tweets written by the crowd. As a result, *Official*-tweets represent a more credible source of information than the other types of tweets for cryptocurrency investors. Therefore, in the following section, we analyze whether the three Twitter samples similarly predict returns.

5.2. Refinements of Twitter-based abnormal attention

As each category of tweets differs in terms of types of authors, functionalities, or reach, we expect that the relationship between expected returns and Twitter-based attention may change depending on the Twitter sample being used. Compared to general tweets, *Mention*-tweets have the particularity of triggering a notification for the recipient of the mention. *Ticker*-tweets, upon being clicked, display the most recent *Ticker*-tweets about the corresponding financial asset.

We now study if the qualitative and quantitative differences between our Twitter samples translate into different relationships in the cross-section of cryptocurrency expected returns. In Table 9, we estimate panel regressions of excess returns on the Twitter samples while controlling for our set of control variables. *Mention*-tweets are positively linked with excess returns in t and $t+1$ at the 1% and 10% significance levels, respectively. In contrast, we do not find any link between *Official*-tweets and returns. The regression coefficients of *Ticker*-tweets are positive and statistically significant at the 1% significance level in both models (1) and (2). The results of both panels are similar.

The results of Table 9 illustrate that the return predictability of Twitter activity is mainly derived from user-generated content and not from the announcements made by cryptocurrencies. This evidence provides additional support for a behavioral-based explanation of our results, as we would expect *Official*-tweets to be associated with returns if our results were driven by a technological innovation channel (Cong et al., 2021). Lastly, we note that *Ticker*-tweets are more strongly related to excess returns than *Mention*-tweets. As the reason behind this difference in predictability power is not clear, we identify two non-exclusive candidate explanations. *Ticker*-tweets could predict returns better because they are more visible and therefore reach more users than *Mention*-tweets. A relatively larger number of users reached would correspond to larger (future) returns in both an overreaction channel (Barber and Odean, 2008; Da et al., 2011) and a network growth channel (Sockin and Xiong, 2023; Cong et al., 2021). Alternatively, investors who use Twitter to inform their investment decisions could choose to primarily consume *Ticker*-tweets as their content is geared toward cryptocurrencies' financial characteristics. The idea that *Ticker*-tweets predict returns because they are more representative of investor attention would be more consistent with an overreaction channel. Our rationale is that the attention of all types of users should matter in a network growth channel and not just investors' attention.

5.3. Twitter textual content and investor attention

In this section, we examine why user-generated content-based attention is able to forecast returns in the cross-section of expected returns, whereas cryptocurrency-generated content does not show predictability. As both *Mention*-tweets and *Ticker*-tweets are able to predict returns despite their different visibility, it seems that investors are concerned about the tweets' content. To test this hypothesis, we now investigate the relationship between the lexicon score computed over the aggregated sample of tweets and daily excess cryptocurrency returns. We also control for textual sentiment to verify that abnormal attention doesn't predict returns because it originates from positive fundamental news. In addition, we also include an interaction term to test whether the return predictability is stronger when an attention-grabbing tweet is posted. We report the regression results in Table 10.

As displayed in Table 8, the regression coefficients of $Abn\,Attention_{i,t}^{All}$ are statistically significant in models (1) and (2) at the 1% significance level. $Sentiment_{i,t}^{All}$, $Lexicon_{i,t}^{All}$, and $Popular\,Tweet_{i,t}^{All}$ are also positively linked with excess returns at time t and $t+1$ at the 1% significance level. The interaction term between Twitter activity and popular tweets is positively connected with returns in (1) and (2) at the 1% and 5% significance levels, respectively.

We find $Sentiment_{i,t}^{All}$ to be positively associated with contemporaneous and next-day returns, which is in line with previous literature on Twitter sentiment (Gu and Kurov, 2020; Jiang, Liu, Roch and Zhou, 2023). The effect of popular users posting on (future) returns is consistent with Benetton et al. (2024); Pedersen (2022), who document that influential users have a strong impact on the investment behavior of their followers and on asset returns. Compared to Benetton et al. (2024), which focus on a set of 75 celebrities, our results show

[23] A risk-based explanation of our results is also possible. Andrei and Hasler (2015) find that variation in investor attention affects the volatility through its impact on the speed of information incorporation into asset prices. Therefore, variation in attention is compensated with similar risk premia variation. We provide additional tests in the appendix showing that our results are best explained by behavioral reasons rather than risk-based mechanisms.

Table 8
Investor attention and cryptocurrency returns, 2018-2022.
The dependent variables are daily excess returns ($Excess\ Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 165 (103) cryptocurrencies. Control variables are defined in Panel A of Table 3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	(2)	(3)	(4)
			$Daily\ Excess\ Return_i$	
Panel A: All cryptocurrencies	t	$t+1$	$t+2$	$t+3$
$Abn\ Attention^{All}_{i,t}$	1.06*** (18.57)	0.16*** (4.29)	0.01 (0.37)	0.02 (0.89)
$Excess\ Return_{i,t}$		−18.94*** (−16.85)	−2.65*** (−5.26)	−0.19 (−0.45)
$Excess\ Return_{i,t-1}$	−18.55*** (−16.46)	−6.11*** (−9.14)	−0.62 (−1.5)	−0.46 (−0.89)
$Excess\ Return_{i,t-2}$	−5.28*** (−7.66)	−1.93*** (−4.04)	−0.79 (−1.39)	−0.25 (−0.57)
$Excess\ Return_{i,t-3}$	−1.65*** (−3.77)	−1.04** (−2.06)	−0.18 (−0.42)	0.71 (1.45)
$Beta_{i,t}$	0.04 (0.34)	0.04 (0.34)	0.15 (1.18)	0.08 (0.63)
$Size_{i,t}$	−0.34*** (−10.51)	−0.3*** (−10.22)	−0.23*** (−8.99)	−0.21*** (−8.29)
$Momentum_{i,t}$	0 (0.01)	−0.06 (−0.84)	−0.08 (−1.4)	−0.13** (−2.11)
$Volatility_{i,t}$	0.42 (0.04)	0.14 (0.01)	−3.21 (−0.55)	−0.87 (−0.13)
$IdioVola_{i,t}$	−7.73 (−1.6)	−7.35 (−1.58)	−2.96 (−0.64)	−4.65 (−1.03)
$Max\ Ret_{i,t}$	2.92 (0.63)	2.63 (0.71)	2.1 (1.36)	2.41 (1.16)
$\Delta Volume_{i,t}$	1.81*** (12.43)	0.26*** (6.7)	−0.04 (−1.18)	−0.01 (−0.32)
$Volume\ Vola_{i,t}$	−0.06 (−0.91)	−0.08 (−1.13)	−0.08 (−1.4)	−0.05 (−0.78)
$Illiquidity_{i,t}$	−0.45 (−0.31)	1.37 (0.9)	1.89 (1.53)	1.3 (1.15)
$Skewness_{i,t}$	0.02 (0.22)	0.01 (0.09)	0.02 (0.42)	0.02 (0.47)
$Kurtosis_{i,t}$	−0.01 (−0.48)	−0.01 (−0.63)	−0.01 (−1.19)	−0.01 (−1.36)
$Co\text{-}Skewness_{i,t}$	−0.03 (−0.35)	−0.01 (−0.14)	0.01 (0.14)	−0.04 (−0.58)
$Co\text{-}Kurtosis_{i,t}$	0.02 (0.63)	0.02 (0.63)	0.01 (0.54)	0.02 (0.74)
$VaR_{i,t}$	−2.24 (−1.25)	−1.49 (−1.02)	0.67 (0.45)	1.32 (0.78)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	255684	254850	254431	254206
Panel B: Filtered ticker				
$Abn\ Attention^{All}_{i,t}$	1.08*** (15.58)	0.14*** (2.96)	0.03 (1.13)	0.04 (1.21)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	160712	160246	159991	159846

that even small influencers can have a significant impact on asset prices. When we interact Twitter abnormal attention with $PopularTweet^{All}_{i,t}$, the effect of Twitter activity on returns is even stronger. This finding also echoes with the results of Hillert et al. (2014), who find that media coverage can amplify behavioral biases. In the context of our study, any users with sufficient popularity on social media seem to be able to exacerbate behavioral biases through their influence on their followers. This latter observation provides more support for the overreaction interpretation of our results.

The regression coefficients of $Abn\ Attention^{All}_{i,t}$ are robust to the inclusion of the new independent variables. We interpret this result as indicative that the number of tweets posted is not subsumed by the textual content nor the sentiment of tweets. This provides support to our explanation that *Ticker*-tweets predict better returns than the other Twitter samples due to their better visibility. Given that lexicon score also forecasts returns, we conclude that both tweet volume and textual content matter in grabbing users' attention. In addition, the positive predictability of the lexicon on (future) returns casts doubts on the interpretation that abnormal attention

Table 9
Refined investor attention and cryptocurrency returns, 2018-2022.
The dependent variables are daily excess returns ($Excess\ Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	(2)	(3)	(4)
		$Daily\ Excess\ Return_i$		
Panel A: All cryptocurrencies	t	$t+1$	$t+2$	$t+3$
$Abn\ Attention^{Mention}_{i,t}$	0.31*** (10.85)	0.04* (1.92)	0.01 (0.32)	−0.02 (−1.3)
$Abn\ Attention^{Official}_{i,t}$	−0.02 (−0.42)	0.02 (0.67)	−0.02 (−0.61)	0.04 (1.22)
$Abn\ Attention^{Ticker}_{i,t}$	1.11*** (16.9)	0.17*** (4.35)	0 (0.16)	0.03 (0.92)
$Excess\ Return_{i,t}$		−18.86*** (−16.09)	−2.44*** (−4.77)	−0.11 (−0.27)
$Excess\ Return_{i,t-1}$	−18.82*** (−16.22)	−5.94*** (−8.55)	−0.55 (−1.3)	−0.58 (−1.07)
$Excess\ Return_{i,t-2}$	−5.3*** (−7.44)	−1.93*** (−3.95)	−0.93 (−1.56)	−0.24 (−0.52)
$Excess\ Return_{i,t-3}$	−1.72*** (−3.85)	−1.13** (−2.15)	−0.12 (−0.26)	0.55 (1.08)
$Beta_{i,t}$	0.07 (0.55)	0.05 (0.42)	0.16 (1.21)	0.09 (0.67)
$Size_{i,t}$	−0.35*** (−10.56)	−0.31*** (−10.04)	−0.24*** (−9.08)	−0.23*** (−8.29)
$Momentum_{i,t}$	0 (−0.02)	−0.05 (−0.76)	−0.08 (−1.35)	−0.12** (−2.05)
$Volatility_{i,t}$	−2.21 (−0.18)	−0.57 (−0.06)	−2.98 (−0.49)	−0.91 (−0.14)
$IdioVola_{i,t}$	−7.09 (−1.44)	−7.17 (−1.51)	−2.72 (−0.58)	−4.2 (−0.91)
$Max\ Ret_{i,t}$	3.74 (0.78)	2.99 (0.78)	2.15 (1.36)	2.5 (1.17)
$\Delta Volume_{i,t}$	1.8*** (11.85)	0.26*** (6.69)	−0.03 (−1.06)	−0.01 (−0.4)
$Volume\ Vola_{i,t}$	−0.06 (−0.88)	−0.07 (−0.94)	−0.08 (−1.34)	−0.05 (−0.83)
$Illiquidity_{i,t}$	−1.23 (−0.68)	0.57 (0.32)	1.2 (0.76)	0.79 (0.58)
$Skewness_{i,t}$	0 (0)	−0.01 (−0.09)	0.01 (0.28)	0.02 (0.4)
$Kurtosis_{i,t}$	0 (−0.3)	−0.01 (−0.53)	−0.01 (−1.24)	−0.01 (−1.39)
$Co\text{-}Skewness_{i,t}$	0 (0)	0.01 (0.16)	0.02 (0.3)	−0.01 (−0.18)
$Co\text{-}Kurtosis_{i,t}$	0.01 (0.34)	0.01 (0.4)	0.01 (0.35)	0.01 (0.68)
$VaR_{i,t}$	−2.57 (−1.37)	−1.53 (−1.02)	0.94 (0.62)	1.71 (1)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243061
Panel B: Filtered ticker				
$Abn\ Attention^{Mention}_{i,t}$	0.32*** (9.84)	0.02 (0.76)	0 (0.03)	−0.01 (−0.24)
$Abn\ Attention^{Official}_{i,t}$	−0.01 (−0.22)	0.04 (0.93)	−0.01 (−0.41)	0 (0.13)
$Abn\ Attention^{Ticker}_{i,t}$	1.12*** (14.27)	0.17*** (3.56)	0.03 (1.06)	0.04 (1.06)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

predicts returns because of its relationship with future user base's growth (Cong et al., 2021; Sockin and Xiong, 2023). In fact, the lexicon loads negatively on tokens which we expect to be frequently used by newcomers, such as terms related to assistance or troubleshooting.

5.4. Technological innovations and investor attention

In this section, we test whether Twitter-based attention predicts expected returns through a technological innovation

Table 10
Investor attention, sentiment and cryptocurrency returns, 2018-2022.
The dependent variables are daily excess returns ($Excess\ Return_i$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 165 (103) cryptocurrencies. Control variables are defined in Panel A of Table 3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	Daily Excess Return$_i$			
Panel A: All cryptocurrencies	t	$t+1$	$t+2$	$t+3$
$Abn\ Attention^{All}_{i,t}$	0.82***	0.11***	0.01	0.04
	(17.02)	(2.92)	(0.52)	(1.42)
$Sentiment^{All}_{i,t}$	3.84***	1.25***	0.28	−0.56*
	(9.69)	(4.07)	(0.91)	(−1.73)
$Abn\ Attention^{All}_{i,t}\ X$	1.39***	0.19**	−0.06	−0.05
$Popular\ Tweet^{All}_{i,t}$	(8.92)	(2.19)	(−0.86)	(−0.83)
$Popular\ Tweet^{All}_{i,t}$	0.32***	0.13***	0.02	−0.02
	(5.24)	(2.68)	(0.44)	(−0.57)
$Lexicon^{All}_{i,t}$	2.15***	0.41***	0.06	−0.13
	(13.92)	(3.27)	(0.56)	(−1.06)
$Excess\ Return_{i,t}$		−19.02***	−2.65***	−0.16
		(−16.96)	(−5.24)	(−0.39)
$Excess\ Return_{i,t-1}$	−18.89***	−6.18***	−0.62	−0.44
	(−17.08)	(−9.26)	(−1.48)	(−0.84)
$Excess\ Return_{i,t-2}$	−5.53***	−1.98***	−0.79	−0.24
	(−8.15)	(−4.14)	(−1.38)	(−0.54)
$Excess\ Return_{i,t-3}$	−1.8***	−1.07**	−0.18	0.72
	(−4.17)	(−2.11)	(−0.41)	(1.46)
$Beta_{i,t}$	0.02	0.04	0.15	0.08
	(0.14)	(0.32)	(1.18)	(0.63)
$Size_{i,t}$	−0.36***	−0.31***	−0.23***	−0.21***
	(−10.55)	(−10.43)	(−9.01)	(−8.04)
$Momentum_{i,t}$	−0.02	−0.06	−0.08	−0.12**
	(−0.31)	(−0.93)	(−1.41)	(−2.07)
$Volatility_{i,t}$	1.51	0.24	−3.24	−0.92
	(0.13)	(0.02)	(−0.55)	(−0.14)
$IdioVola_{i,t}$	−9.53**	−7.66*	−2.95	−4.52
	(−1.96)	(−1.65)	(−0.64)	(−1)
$Max\ Ret_{i,t}$	3.14	2.68	2.1	2.39
	(0.68)	(0.72)	(1.36)	(1.15)
$\Delta Volume_{i,t}$	1.78***	0.25***	−0.04	−0.01
	(12.44)	(6.62)	(−1.19)	(−0.29)
$Volume\ Vola_{i,t}$	−0.07	−0.08	−0.08	−0.05
	(−1.04)	(−1.17)	(−1.41)	(−0.78)
$Illiquidity_{i,t}$	−0.38	1.32	1.87	1.33
	(−0.27)	(0.87)	(1.52)	(1.17)
$Skewness_{i,t}$	0.02	0.01	0.02	0.02
	(0.2)	(0.09)	(0.42)	(0.47)
$Kurtosis_{i,t}$	−0.01	−0.01	−0.01	−0.01
	(−0.55)	(−0.64)	(−1.18)	(−1.36)
$Co\text{-}Skewness_{i,t}$	−0.02	−0.01	0.01	−0.05
	(−0.24)	(−0.13)	(0.15)	(−0.59)
$Co\text{-}Kurtosis_{i,t}$	0.02	0.02	0.01	0.02
	(0.65)	(0.67)	(0.56)	(0.72)
$VaR_{i,t}$	−2.42	−1.56	0.66	1.35
	(−1.35)	(−1.07)	(0.44)	(0.8)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	255684	254850	254431	254206
Panel B: Filtered ticker				
$Abn\ Attention^{All}_{i,t}$	0.82***	0.08*	0.04	0.06
	(14.21)	(1.76)	(1.21)	(1.63)
$Sentiment^{All}_{i,t}$	3.73***	1.22***	0.08	−0.72*
	(7.95)	(3.17)	(0.22)	(−1.82)
$Abn\ Attention^{All}_{i,t}\ X$	1.54***	0.24**	−0.07	−0.04
$Popular\ Tweet^{All}_{i,t}$	(8.23)	(2.2)	(−0.82)	(−0.64)
$Popular\ Tweet^{All}_{i,t}$	0.32***	0.14***	0.04	−0.04
	(5.17)	(2.83)	(0.94)	(−0.61)
$Lexicon^{All}_{i,t}$	2.19***	0.43***	0.11	−0.09
	(11.26)	(2.59)	(0.87)	(−0.63)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	160712	160246	159991	159846

channel. Twitter content could potentially forecast future returns through its predictability of technological improvement in the blockchain. This channel is plausible as social media is frequently used by developers to exchange ideas about potential improvements or to signal cybersecurity breaches. In the literature, Cong et al. (2021) theoretically show that cryptocurrencies valuations are influenced by technological improvements. Empirically, technological innovation is linked negatively to delisting probability (Liu et al., 2022a) and positively to ICO success (Lyandres et al., 2022).

To proxy for technological improvements, we utilize the log-difference of the number of commits published for each date and cryptocurrency on GitHub + 1. We believe that this proxy for technological improvements is appropriate because commits capture all code revisions made by developers. Therefore, any new feature or improvement made on the underlying technology used by a specific cryptocurrency will be reflected in its commit history.

Regression results are reported in Table 11. We find that contemporaneous and future cryptocurrency's technological development is strongly associated with *Official*-tweets. The contemporaneous relationship is positive and statistically significant at the 1% level. The regression coefficients of *Official*-tweets in the other models are all negative and statistically significant at the 1% or 10% level. We do not observe any links between logarithmic change in daily commit and the other Twitter samples. We interpret the results as indicative that *Ticker*-tweets and *Mention*-tweets do not predict returns through a technology innovation channel. In contrast, the significant association between the number of *Official*-tweets and technological innovation further confirms that tweets posted by cryptocurrencies can be interpreted as news.

6. Additional Empirical Tests

6.1. Robustness

To provide robustness of our results between abnormal attention and future cryptocurrency returns, we estimate panel regression models with slight modifications compared to Table 8. The robustness checks are reported in Table 12. In column one, we consider raw attention in the regression setup which is defined as:

$$Raw\ Attention_{i,t} = Log(1 + Number\ Of\ Tweets_{i,t}) \quad (3)$$

In the other columns, we use abnormal attention as our main independent variable as in Table 8. In models (2) and (3), we filter the sample to keep only cryptocurrencies classified as coins or as tokens, respectively, to verify that our results are not driven by characteristic differences between coins and tokens. We also check if our results are influenced by the rally of meme stocks led by R/WallStreetBets in 2021 which also affected some cryptocurrencies. For this purpose, we restrict our sample in models (4) and (5) to observations occurring before and after the first three months of 2021. In models (6) and (7), we explore if our results hold for both large and small assets by restricting our sample to assets with a market capitalization, respectively, above or below the median market capitalization. Finally, model (8) excludes pump and dump events to verify that our results are not driven by deliberate price manipulation schemes. We use data from Ardia and Bluteau (2024) to identify and filter out pump and dump events from our sample.[24] The pump and dump data lists both successful and unsuccessful events. By conservatism, we choose to remove the 3237 cryptocurrency-weeks in our sample concerned by such an event.

In all model specifications of Panel A, the regression coefficients of Twitter attention are significantly positive which is consistent with our main model specification presented in Table 8. When regressing expected returns at $t + 1$ in Panel B, we see that the regression coefficients are statistically significant for all specifications, except when small coins are removed from the sample. This observation is consistent with a limits-to-arbitrage explanation, as larger cryptocurrencies are easier to short-sell than smaller assets.

6.2. Additional tests and alternative explanation of the results

In this section, we conduct complementary tests to better characterize the empirical links between Twitter abnormal attention and cryptocurrency market variables.

To provide additional evidence that our results are indeed driven by an overreaction narrative, we investigate how Twitter-based attention predicts contemporaneous and future change in trading volume in a panel regression setting. Results are reported in Table 13. *Mention*-tweets and *Ticker*-tweets are both strongly linked with contemporaneous and future change in volume up to $t + 3$. Both types of tweets are positively linked with contemporaneous volume at the 1% significance level and negatively linked with future trading volume. The signs of the regression coefficients of *Ticker*-tweets and *Mention*-tweets make intuitive sense, as an overreaction channel is characterized by an increased buying pressure that decreases over the subsequent days.

To verify that the absence of reversal is plausibly due to limits-to-arbitrage (Schmeling et al., 2023), we investigate the relationship between abnormal attention and cryptocurrency squared returns. Schmeling et al. (2023) argue that due to cryptocurrency futures exchanges rules about maximum losses on futures positions, even small price fluctuations can easily trigger the liquidation of an entire future position. This special set of rules impedes the ability of sophisticated investors to implement common arbitrage strategies, as highlighted by Schmeling et al. (2023). To align with our results, rises in abnormal attention need to predict increases in volatility to discourage sophisticated investors from arbitraging away the short term positive price pressure. This is indeed what we find in Table 14. Abnormal attention based on *Ticker*-tweets strongly predicts future squared returns up to $t+4$ making it risky for arbitrageurs to bet on price decreases. Furthermore, price increases seem to be driven by behavioral factors.

[24] The data has been made publicly available by the authors at https://doi.org/10.5281/zenodo.12019080

Table 11
Refined investor attention and GitHub commits.
The dependent variables are daily log-changes in the number of commits ($\Delta Commit_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 136 (86) cryptocurrencies. Control variables are defined in Panel A of Table 3. Standard errors are clustered along weeks and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
			$\Delta\, Commit_i$	
Panel A: All cryptocurrencies	t	$t+1$	$t+2$	$t+3$
$Abn\ Attention^{Mention}_{i,t}$	−0.33	−0.41	−0.08	−0.12
	(−1.07)	(−1.04)	(−0.24)	(−0.71)
$Abn\ Attention^{Official}_{i,t}$	5.61***	−6.62***	−5.71***	−0.81*
	(5.37)	(−6.5)	(−5.28)	(−1.7)
$Abn\ Attention^{Ticker}_{i,t}$	0.06	−0.04	−0.49	0.13
	(0.16)	(−0.12)	(−1.43)	(0.59)
$Excess\ Return_{i,t}$	1.84	−0.6	0.93	1.95
	(0.8)	(−0.25)	(0.43)	(0.68)
$Excess\ Return_{i,t-1}$	−1.71	0.05	3.45	−3.33
	(−0.75)	(0.02)	(1.27)	(−1.59)
$Excess\ Return_{i,t-2}$	−1.26	3.69	−3.8*	−0.34
	(−0.58)	(1.36)	(−1.75)	(−0.15)
$Excess\ Return_{i,t-3}$	2.91	−3.3	−0.73	0.98
	(1.06)	(−1.55)	(−0.3)	(0.48)
$Beta_{i,t}$	0.17	0.13	−0.01	−0.19
	(0.27)	(0.2)	(−0.02)	(−0.29)
$Size_{i,t}$	0.07	−0.19	−0.14	−0.07
	(0.23)	(−0.59)	(−0.45)	(−0.23)
$Momentum_{i,t}$	−0.07	0.19	0.26	0.06
	(−0.2)	(0.71)	(0.9)	(0.23)
$Volatility_{i,t}$	1.83	−4.38	−1.2	11.64
	(0.06)	(−0.14)	(−0.04)	(0.35)
$IdioVola_{i,t}$	8.54	3.06	−10.48	−17.62
	(0.29)	(0.11)	(−0.36)	(−0.59)
$Max\ Ret_{i,t}$	−3	−2.42	1.24	0.31
	(−1.26)	(−0.96)	(0.6)	(0.1)
$\Delta Volume_{i,t}$	0.34	0.31	−0.54*	0.16
	(0.98)	(0.95)	(−1.77)	(0.64)
$Volume\ Vola_{i,t}$	−0.21	−0.09	−0.15	−0.15
	(−1.09)	(−0.44)	(−0.73)	(−0.75)
$Illiquidity_{i,t}$	4.73	2.19	−2.11	−4.26
	(0.72)	(0.36)	(−0.32)	(−0.66)
$Skewness_{i,t}$	−0.04	−0.02	−0.09	−0.01
	(−0.27)	(−0.19)	(−0.65)	(−0.07)
$Kurtosis_{i,t}$	−0.02	0.01	0.02	0.02
	(−0.33)	(0.2)	(0.42)	(0.61)
$Co\text{-}Skewness_{i,t}$	0.34	0.56**	0.42	0.21
	(0.85)	(1.97)	(1.25)	(0.81)
$Co\text{-}Kurtosis_{i,t}$	−0.02	−0.04	−0.04	−0.04
	(−0.11)	(−0.23)	(−0.22)	(−0.23)
$VaR_{i,t}$	2.28	−2.68	−5.37	−4.34
	(0.55)	(−0.56)	(−1.25)	(−1)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	218873	218769	218668	218568
Panel B: Filtered ticker				
$Abn\ Attention^{Mention}_{i,t}$	−0.34	−0.93*	−0.21	−0.12
	(−1.02)	(−1.69)	(−0.47)	(−0.48)
$Abn\ Attention^{Official}_{i,t}$	6.56***	−7.47***	−6.78***	−1.14*
	(5.06)	(−5.9)	(−4.62)	(−1.79)
$Abn\ Attention^{Ticker}_{i,t}$	−0.13	0.14	−0.6*	0.19
	(−0.34)	(0.38)	(−1.67)	(0.63)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	139313	139244	139178	139113

Our results could also be explained by the model Andrei and Hasler (2015), who document similar relationships between attention, expected returns, and volatility as in this paper. The authors argue and show that increases in investor attention accelerate the incorporation of new information into asset prices, which leads to stronger price variations and risk premia. To shed light on whether our results indeed relate to Andrei and Hasler (2015), we investigate how abnormal attention relates to idiosyncratic squared returns. Idiosyncratic squared returns have the advantage of capturing large changes in returns that are not driven by variation in risk factors. The results are reported in Table 15.

Table 12
Robustness checks.
The dependent variables are daily excess returns ($Excess\ Return_t$). The regression spans the time period 2018 to 2022 for a sample of 165 cryptocurrencies. Control variables are defined in Panel A of Table 3. The first model (1) uses raw attention instead of abnormal change as the main independent variable. Models (2) and (3) restrict the sample to coins only or tokens only, respectively. Models (4) and (5) restrict the sample respectively to observations occurring before or after the GameStop short-squeeze. For models (6) and (7), we restrict the sample to assets with a market cap below or above the median, respectively. Finally, model (8) excludes pump and dump events from the sample. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Daily\ Excess\ Return_{i,t}$							
	Raw Attention	*Coins Only*	*Tokens Only*	*Before GameStop*	*After GameStop*	*Large Only*	*Small Only*	*No P&D Events*
Panel A: Contemporaneous relationship								
$Abn\ Attention^{All}_{i,t}$	0.58***	1.02***	1.09***	1.11***	0.88***	1.49***	0.62***	1.05***
	(10.29)	(14.18)	(12.25)	(17.19)	(11.31)	(19.41)	(11.64)	(18.38)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	255684	147494	108190	165359	79856	129704	125160	252490
	$Daily\ Excess\ Return_{i,t+1}$							
Panel B: Predictive relationship								
$Abn\ Attention^{All}_{i,t}$	0.11***	0.15***	0.15***	0.17***	0.13**	0.03	0.13***	0.17***
	(3.68)	(3.19)	(2.58)	(3.84)	(2.41)	(0.68)	(2.89)	(4.42)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	254850	147023	107827	164984	79440	129570	124461	251659

We document a strong relationship between daily idiosyncratic squared returns and abnormal attention measured on *Mention*-tweets and *Ticker*-tweets. The regression coefficients are statistically significant at the 5% or 1% in all model specifications. As the dependent variable accounts for risk factor variations, Twitter attention's relationship with expected returns is not purely driven by variation in priced risk. This finding provides support for the overreaction channel to be the most likely mechanism behind our results.

7. Conclusion

Given the size of the cryptocurrency market and the importance of Twitter as a source of information, studying their interplay is essential for understanding the cross-section of cryptocurrency returns. In the literature, the evidence on the link between Twitter activity and cryptocurrency prices is mixed. Consistent with Benedetti and Kostovetsky (2021), we show that Twitter impacts the cross-section of cryptocurrency expected returns through an overreaction channel (Barber and Odean, 2008; Da et al., 2011). However, we do not document any price reversals, which is plausibly due to limits-to-arbitrage, as already noted by (Schmeling et al., 2023). We further find that the tweeting activity of popular users exacerbates behavioral biases and therefore the overreaction effect (Hillert et al., 2014). We interpret this evidence as a warning sign about the ability of influential users to manipulate asset prices.

To better understand the relationship of Twitter attention with expected returns, we refine our abnormal attention by utilizing different samples of tweets. We show that the return predictability power of Twitter activity mainly arises from *Ticker*-tweets. In contrast, we find no association between tweets posted by official cryptocurrency channels and future returns, despite *Official*-tweets being able to forecast future innovations in the implementation code of each cryptocurrency (unlike the other Twitter samples). This empirical finding does not align with the predictions of the theoretical model of Cong et al. (2021) that cryptocurrency valuations should be linked with technological improvements. Lastly, we document that the return predictability of *Ticker*-tweets is partly due to their salience and unique textual content, which caters to the preferences of retail investors. Overall, our results emphasize the heterogeneity of social media content, highlighting the need for researchers and practitioners to carefully consider which types of social media content best suit their needs.

CRediT authorship contribution statement

Arnaud T. Maître: Writing- review and editing, Writing- original draft, Methodology, Investigation, Formal analysis, Conceptualization.. **Nikolay Pugachyov :** Writing- review and editing, Writing- original draft, Methodology, Investigation, Formal analysis, Conceptualization.. **Florian Weigert :** Writing- review and editing, Writing- original draft, Methodology, Investigation, Formal analysis, Conceptualization..

Table 13
Refined investor attention and changes in volume.
The dependent variables are daily logarithmic changes in trading volume. The regression spans the period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 3. Standard errors are clustered along weeks and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	$Daily\,\Delta\,Volume_i$			
Panel A: All cryptocurrencies	t	$t+1$	$t+2$	$t+3$
$Abn\ Attention^{Mention}_{i,t}$	1.26***	−1.35***	−0.73***	−0.23
	(5.99)	(−7.23)	(−3.8)	(−1.52)
$Abn\ Attention^{Official}_{i,t}$	0.31	0.84***	−0.54*	0.47*
	(1.03)	(2.89)	(−1.75)	(1.73)
$Abn\ Attention^{Ticker}_{i,t}$	7.31***	−3.42***	−1.8***	−0.65***
	(13.03)	(−9.97)	(−6.96)	(−3.34)
$Excess\ Return_{i,t}$	169.03***	44.07***	−46.82***	−13.45***
	(10.53)	(9.95)	(−9.64)	(−4.44)
$Excess\ Return_{i,t-1}$	3.93	−57.62***	−26.09***	−5.53*
	(0.86)	(−9.25)	(−8.9)	(−1.67)
$Excess\ Return_{i,t-2}$	−60.01***	−46.92***	−12.48***	−7.42***
	(−9.76)	(−12.07)	(−3.32)	(−3)
$Excess\ Return_{i,t-3}$	−29.21***	−14.8***	−6.87***	−1.37
	(−9.31)	(−4.8)	(−2.67)	(−0.51)
$Beta_{i,t}$	−1.34***	−1.79***	−0.63	−0.87**
	(−2.94)	(−3.01)	(−1.34)	(−2.28)
$Size_{i,t}$	−0.2	−0.79***	−0.71***	−0.4***
	(−1.54)	(−5.28)	(−5.06)	(−3.3)
$Momentum_{i,t}$	−1.28***	−1.14***	−0.56***	−0.56***
	(−5.28)	(−4.26)	(−2.77)	(−3.11)
$Volatility_{i,t}$	44.57	75.76**	26.27	21.26
	(1.31)	(2.21)	(0.91)	(0.77)
$Idio\,Vola_{i,t}$	−82.84***	−113.26***	−47.35**	−39.21**
	(−4.09)	(−4.75)	(−2.39)	(−2.45)
$Max\ Ret_{i,t}$	12.13	7.97	1.78	1.53
	(1.04)	(0.83)	(0.28)	(0.21)
$\Delta Volume_{i,t}$		−34.06***	−8.12***	−1.92***
		(−67.28)	(−19.27)	(−4.22)
$Volume\,Vola_{i,t}$	0.89**	0.98**	0.49	0.49
	(2.36)	(2.27)	(1.49)	(1.55)
$Illiquidity_{i,t}$	35.42***	36.73***	20.1**	−0.17
	(3.47)	(2.82)	(2)	(−0.02)
$Skewness_{i,t}$	−0.09	0.1	0.14	0.2
	(−0.35)	(0.37)	(0.74)	(1.03)
$Kurtosis_{i,t}$	−0.07*	−0.14***	−0.08**	−0.07*
	(−1.69)	(−3.08)	(−2.25)	(−1.74)
$Co\text{-}Skewness_{i,t}$	0.29	0.48	0.4	−0.13
	(0.82)	(1.05)	(1.12)	(−0.4)
$Co\text{-}Kurtosis_{i,t}$	0.13	0.33**	0.22	0.17
	(1.08)	(2)	(1.53)	(1.53)
$VaR_{i,t}$	7.08	2.76	−3.95	−6.53
	(1.16)	(0.3)	(−0.66)	(−1.5)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243970	243571	243362
Panel B: Filtered ticker				
$Abn\ Attention^{Mention}_{i,t}$	1.16***	−1.62***	−0.55**	−0.34*
	(4.3)	(−7)	(−2.26)	(−1.72)
$Abn\ Attention^{Official}_{i,t}$	0.39	1***	−0.75**	0.81**
	(1.02)	(2.59)	(−2.08)	(2.32)
$Abn\ Attention^{Ticker}_{i,t}$	7.48***	−3.91***	−1.7***	−0.57**
	(10.75)	(−9.31)	(−5.32)	(−2.45)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156868	156616	156475

References

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Ammann, M., Burdorf, T., Liebi, L., Stöckl, S., 2022. Survivorship and delisting bias in cryptocurrency markets. Unpublished Working Paper.

Andrei, D., Hasler, M., 2015. Investor attention and stock market volatility. Review of Financial Studies 28, 33–72.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259–299.

Ardia, D., Bluteau, K., 2024. Twitter and cryptocurrency pump-and-dumps. International Review of Financial Analysis 95, 103479.

Table 14
Refined investor attention and squared returns.
The dependent variables are daily squared excess returns. The regression spans the time period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 3. Standard errors are clustered along week and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	Daily Squared R_i^e			
Panel A: All cryptocurrencies		$t+1$	$t+2$	$t+3$
$Abn\ Attention_{i,t}^{Mention}$	0.06***	0.03***	0.02**	0.01
	(6.28)	(3.49)	(2.26)	(1.5)
$Abn\ Attention_{i,t}^{Official}$	−0.01	−0.02*	−0.01	0.02
	(−1.16)	(−1.73)	(−0.77)	(1.24)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.09***	0.08***	0.06***
	(8.35)	(5.68)	(4.8)	(3.44)
$Excess\ Return_{i,t}$	12.98***	0.28	0.58***	0.34
	(17.16)	(0.87)	(2.96)	(1.54)
$Excess\ Return_{i,t-1}$	2.84***	0.7***	0.51**	0.34
	(9.01)	(3.13)	(2.09)	(1.21)
$Excess\ Return_{i,t-2}$	1.6***	0.59**	0.36	0.21
	(7.67)	(2.5)	(1.37)	(1.44)
$Excess\ Return_{i,t-3}$	0.74***	0.18	0.04	0.22
	(4.05)	(0.9)	(0.29)	(1.23)
$Beta_{i,t}$	−0.21***	−0.23***	−0.23***	−0.21***
	(−2.91)	(−3)	(−2.8)	(−2.83)
$Size_{i,t}$	0.01	−0.05***	−0.05***	−0.06***
	(0.67)	(−3.87)	(−3.43)	(−4.04)
$Momentum_{i,t}$	0.07**	0.07**	0.02	0.01
	(2.33)	(2.29)	(0.65)	(0.29)
$Volatility_{i,t}$	16.71***	16.28***	11.03**	9.46***
	(4.62)	(4.86)	(2.5)	(2.66)
$IdioVola_{i,t}$	9.66***	6.39***	5.26**	6***
	(4.18)	(2.94)	(2.37)	(2.96)
$Max\ Ret_{i,t}$	−6.31***	−5.08***	−2.68*	−2.56**
	(−4.65)	(−4.15)	(−1.73)	(−2.34)
$\Delta Volume_{i,t}$	0.24***	0.08***	0.02	−0.01
	(10.35)	(5.77)	(1.44)	(−0.83)
$Volume\ Vola_{i,t}$	0.02	0.04	0.04	0.05
	(0.6)	(0.9)	(0.93)	(1.14)
$Illiquidity_{i,t}$	1	0.04	0.75	0.63
	(0.63)	(0.03)	(0.49)	(0.48)
$Skewness_{i,t}$	0.04	0.05	0	0
	(1.26)	(1.59)	(0.11)	(−0.17)
$Kurtosis_{i,t}$	−0.02***	−0.03***	−0.02***	−0.02***
	(−4.08)	(−5.66)	(−3.63)	(−4.06)
$Co\text{-}Skewness_{i,t}$	0.09*	0.11**	0.09*	0.12**
	(1.85)	(2.35)	(1.92)	(2.52)
$Co\text{-}Kurtosis_{i,t}$	0.06***	0.07***	0.06***	0.06***
	(3.86)	(4.33)	(3.59)	(3.81)
$VaR_{i,t}$	−3.28***	−3.22***	−3.89***	−3.95***
	(−2.59)	(−2.72)	(−2.75)	(−3.06)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243061
Panel B: Filtered ticker				
$Abn\ Attention_{i,t}^{Mention}$	0.05***	0.02	0.02	0.01
	(4.23)	(1.51)	(1.41)	(0.88)
$Abn\ Attention_{i,t}^{Official}$	−0.01	−0.01	−0.01	0.02
	(−1.01)	(−0.58)	(−0.94)	(1.27)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.1***	0.09***	0.06***
	(7.73)	(5.14)	(4.94)	(3.32)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. Journal of Finance 68, 929–985.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., 2018. The economic effects of social networks: Evidence from the housing market. Journal of Political Economy 126, 2224–2276.

Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. Journal of Financial Economics 99, 427–446.

Barber, B.M., Huang, X., Odean, T., Schwarz, C., 2022. Attention-induced trading and returns: Evidence from robinhood users. Journal of Finance 77, 3141–3190.

Table 15
Refined investor attention and idiosyncratic squared returns.
The dependent variables are daily idiosyncratic squared excess returns. The regression spans the time period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table 3. Standard errors are clustered along week and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
			Daily Idiosyncratic $(R_i^e)^2$	
Panel A: All cryptocurrencies	t	$t+1$	$t+2$	$t+3$
$Abn\ Attention_{i,t}^{Mention}$	0.07***	0.03***	0.03***	0.02**
	(6.9)	(4.1)	(2.84)	(2.07)
$Abn\ Attention_{i,t}^{Official}$	−0.02	−0.02**	−0.01	0.02
	(−1.51)	(−1.97)	(−0.72)	(1.47)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.09***	0.08***	0.06***
	(8.32)	(5.57)	(4.85)	(3.55)
$Excess\ Return_{i,t}$	12.45***	0.41	0.62***	0.28
	(17.86)	(1.32)	(3.06)	(1.25)
$Excess\ Return_{i,t-1}$	2.88***	0.81***	0.49**	0.4
	(9.29)	(3.71)	(2.04)	(1.45)
$Excess\ Return_{i,t-2}$	1.65***	0.6***	0.42	0.27*
	(7.89)	(2.58)	(1.6)	(1.72)
$Excess\ Return_{i,t-3}$	0.73***	0.22	0.1	0.27
	(3.91)	(1.08)	(0.63)	(1.57)
$Beta_{i,t}$	−0.27***	−0.28***	−0.29***	−0.27***
	(−3.69)	(−3.74)	(−3.47)	(−3.55)
$Size_{i,t}$	−0.01	−0.07***	−0.07***	−0.08***
	(−0.7)	(−5.24)	(−4.62)	(−5.32)
$Momentum_{i,t}$	0.07**	0.06**	0.01	0
	(2.2)	(2.06)	(0.27)	(−0.1)
$Volatility_{i,t}$	14.65***	14.79***	10.41**	9.29***
	(4.08)	(4.54)	(2.46)	(2.7)
$IdioVola_{i,t}$	11.16***	7.59***	6.42***	7.12***
	(4.9)	(3.46)	(2.88)	(3.39)
$Max\ Ret_{i,t}$	−6.06***	−4.9***	−2.75*	−2.78***
	(−4.56)	(−4.09)	(−1.82)	(−2.6)
$\Delta Volume_{i,t}$	0.23***	0.07***	0.02	−0.01
	(9.97)	(5.43)	(1.33)	(−0.61)
$Volume\ Vola_{i,t}$	0.02	0.04	0.05	0.05
	(0.69)	(1.04)	(1.09)	(1.27)
$Illiquidity_{i,t}$	1.03	−0.12	0.54	0.42
	(0.63)	(−0.07)	(0.36)	(0.32)
$Skewness_{i,t}$	0.04	0.04	0	0
	(1.11)	(1.44)	(0.03)	(−0.11)
$Kurtosis_{i,t}$	−0.02***	−0.03***	−0.02***	−0.02***
	(−4.06)	(−5.71)	(−3.71)	(−4.36)
$Co\text{-}Skewness_{i,t}$	0.11**	0.12***	0.11**	0.14***
	(2.39)	(2.75)	(2.28)	(2.82)
$Co\text{-}Kurtosis_{i,t}$	0.07***	0.08***	0.06***	0.06***
	(4.03)	(4.54)	(3.78)	(4.08)
$VaR_{i,t}$	−3.4***	−3.27***	−3.8***	−3.85***
	(−2.81)	(−2.88)	(−2.71)	(−2.94)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243060
Panel B: Filtered ticker				
$Abn\ Attention_{i,t}^{Mention}$	0.06***	0.02*	0.02*	0.01
	(4.89)	(1.92)	(1.65)	(1.27)
$Abn\ Attention_{i,t}^{Official}$	−0.02*	−0.01	−0.01	0.02
	(−1.72)	(−1.1)	(−0.78)	(1.21)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.1***	0.08***	0.07***
	(7.85)	(4.99)	(4.9)	(3.66)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Barber, B.M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 21, 785–818.

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

Benedetti, H., Kostovetsky, L., 2021. Digital tulips? returns to investors in initial coin offerings. Journal of Corporate Finance 66, 101786.

Benetton, M., Mullins, W., Niessner, M., Toczynski, J., 2024. Celebrity influencers: This is not financial advice. Unpublished Working Paper.

Borri, N., Massacci, D., Rubin, M., Ruzzi, D., 2022. Crypto risk premia. Unpublished Working Paper.

Bose, D., Cordes, H., Nolte, S., Schneider, J.C., Camerer, C.F., 2022. Decision weights for experimental asset prices based on visual salience. Review of Financial Studies 35, 5094–5126.

Broadstock, D.C., Zhang, D., 2019. Social-media and intraday stock returns: The pricing power of sentiment. Finance Research Letters 30, 116–123.

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

Chordia, T., Subrahmanyam, A., Anshuman, V.R., 2001. Trading activity and expected stock returns. Journal of financial Economics 59, 3–32.

Cong, L.W., Li, Y., Wang, N., 2021. Tokenomics: Dynamic adoption and valuation. Review of Financial Studies 34, 1105–1155.

Cookson, J.A., Engelberg, J.E., Mullins, W., 2023. Echo chambers. Review of Financial Studies 36, 450–500.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. Journal of Finance 66, 1461–1499.

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

Fang, H., Lai, T.Y., 1997. Co-kurtosis and capital asset pricing. Financial Review 32, 293–307.

Fong, W.M., Toh, B., 2014. Investor sentiment and the max effect. Journal of Banking & Finance 46, 190–201.

Frydman, C., Wang, B., 2020. The impact of salience on investor behavior: Evidence from a natural experiment. Journal of Finance 75, 229–276.

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

Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.

Hillert, A., Jacobs, H., Müller, S., 2014. Media makes momentum. Review of Financial Studies 27, 3467–3501.

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

Hosseini, A., Jostova, G., Philipov, A., Savickas, R., 2020. The social media risk premium. Unpublished Working Paper.

Jiang, H., Liu, P., Roch, A.F., Zhou, X., 2023. Social media sentiment and bitcoin price dynamics: Evidence from 36 million postings. Unpublished Working Paper.

Jondeau, E., Rockinger, M., 2006. The copula-garch model of conditional dependencies: An international stock market application. Journal of international money and finance 25, 827–853.

Kraus, A., Litzenberger, R.H., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31, 1085–1100.

Kumar, A., Ruenzi, S., Ungeheuer, M., 2021. Daily winners and losers. Unpublished Working Paper.

Liu, Y., Sheng, J., Wang, W., 2022a. Technology and cryptocurrency valuation: Evidence from machine learning. Unpublished Working Paper.

Liu, Y., Tsyvinski, A., 2021. Risks and returns of cryptocurrency. Review of Financial Studies 34, 2689–2727.

Liu, Y., Tsyvinski, A., Wu, X., 2022b. Common risk factors in cryptocurrency. Journal of Finance 77, 1133–1177.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. Journal of Finance 66, 35–65.

Lyandres, E., Palazzo, B., Rabetti, D., 2022. Initial coin offering (ico) success and post-ico performance. Management Science 68, 8658–8679.

Lyócsa, Š., Baumöhl, E., Vỳrost, T., 2022. Yolo trading: Riding with the herd during the gamestop episode. Finance Research Letters 46, 102359.

Pedersen, L.H., 2022. Game on: Social networks and markets. Journal of Financial Economics 146, 1097–1119.

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

Schmeling, M., Schrimpf, A., Todorov, K., 2023. Crypto carry. Unpublished Working Paper.

Sockin, M., Xiong, W., 2023. A model of cryptocurrencies. Management Science 69, 6684–6707.

Solomon, D.H., Soltes, E., Sosyura, D., 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. Journal of Financial Economics 113, 53–72.

Tan, H.T., Wang, E.Y., Zhou, B., 2015. How does readability influence investors' judgments? consistency of benchmark performance matters. The Accounting Review 90, 371–393.

cfr working paper series

centre for financial research
cologne

cfr working papers are available for download from www.cfr-cologne.de.

2025

No.	Author(s)	Title
25-02	A. T. Maître, N. Pugachyov, F. Weigert	Twitter-Based Attention and the Cross-Section of Cryptocurrency Returns
25-01	N. Käfer, M. Mörke, F. Weigert, T. Wiest	A Bayesian Stochastic Discount Factor for the Cross-Section of Individual Equity Options

2024

No.	Author(s)	Title
24-06	V. Beyer, T. Bauckloh	Non-Standard Errors in Carbon Premia
24-05	C. Achilles, P. Limbach, M. Wolff, A. Yoon	Inside the Blackbox of Firm Environmental Efforts: Evidence from Emissions Reduction Initiatives
24-04	Ivan T. Ivanov, T. Zimmermann	The “Privatization” of Municipal Debt
24-03	T. Dyer, G. Köchling, P. Limbach	Traditional Investment Research and Social Networks: Evidence from Facebook Connections
24-02	A. Y. Chen, A. Lopez-Lira, T. Zimmermann	Does Peer-Reviewed Research Help Predict Stock Returns?
24-01	G. Cici, P. Schuster, F. Weishaupt	Once a Trader, Always a Trader: The Role of Traders in Fund Management

2023

No.	Author(s)	Title
23-08	A. Braun, J. Braun, F. Weigert	Extreme Weather Risk and the Cost of Equity
23-07	A. G. Huang, R. Wermers, J. Xue	“Buy the Rumor, Sell the News”: Liquidity Provision by Bond Funds Following Corporate News Events
23-06	J. Dörries, O. Korn, G. J. Power	How Should the Long-term Investor Harvest Variance Risk Premiums?
23-05	V. Agarwal, W. Jiang, Y. Luo, H. Zou	The Real Effect of Sociopolitical Racial Animus: Mutual Fund Manager Performance During the AAPI Hate
23-04	V. Agarwal, B. Barber, S. Cheng, A. Hameed, H. Shanker, A. Yasuda	Do Investors Overvalue Startups? Evidence from the Junior Stakes of Mutual Funds
23-03	A. Höck, T. Bauckloh, M. Dumrose, C. Klein	ESG Criteria and the Credit Risk of Corporate Bond Portfolios
23-02	T. Bauckloh, J. Dobrick, A. Höck, S. Utz, M. Wagner	In partnership for the goals? The level of agreement between SDG ratings
23-01	F. Simon, S. Weibels, T. Zimmermann	Deep Parametric Portfolio Policies

This document only covers the most recent cfr working papers. A full list can be found at www.cfr-cologne.de.

centre for financial research
cfr/university of cologne
Albertus-Magnus-Platz
D-50923 cologne
fon +49(0)221-470-6995
fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de