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

Arnaud T. Maître, Nikolay Pugachyov, and Florian Weigert*

This Version: February 2024

Abstract

This paper investigates how investor abnormal attention affects the cross-section of cryptocurrency returns in the period from 2018 to 2022. We capture investor abnormal attention by logarithmic changes in the daily amount of Twitter posts on individual cryptocurrencies. Our results reveal that investor abnormal attention is positively associated with contemporaneous returns. However, consistent with an investor overreaction narrative, cryptocurrencies with high abnormal attention earn low future performance after the initial increase. We find empirical evidence that Twitter-based abnormal attention is not only a predictor of future returns, but also forecasts technological innovations on the blockchain.

Keywords: Bitcoin, Cryptocurrencies, Twitter attention, textual sentiment

JEL Classification Numbers: G12, G41

Arnaud T. Maître is from the University of Neuchâtel, Institute of Financial Analysis, Rue Abram-Louis-Breguet 2, 2000 Neuchâtel, Switzerland. Email: arnaud.maitre@unine.ch. Tel: +41-32-718-1918. Nikolay Pugachyov is from the University of Neuchâtel, Institute of Financial Analysis, Rue Abram-Louis-Breguet 2, 2000 Neuchâtel, Switzerland. Email: nikolay.pugachyov@unine.ch. Tel: +41-32-718-1468. Florian Weigert is from the University of Neuchâtel, Institute of Financial Analysis, Rue Abram-Louis-Breguet 2, 2000 Neuchâtel, Switzerland. Email: florian.weigert@unine.ch. Tel: +41-31-718-1331. We thank Twitter for having provided access to their academic API, seminar participants at the University of Neuchâtel IAF research day 2023 as well as conference participants at the 2nd International Fintech Conference at Cardiff Business School and the 7th Shanghai-Edinburgh-UCL Fintech Conference for helpful comments. Florian Weigert would like to thank the Centre for Financial Research (CFR) in Cologne for their support. All errors are our own.

1 Introduction

Social media has significant effects on financial markets, as illustrated by the involvement of WallStreetBets subreddit in the Gamestop short-squeeze or famous "to the moon tweets" with subsequent Dogecoin price hikes. Social media platforms are particularly important for cryptocurrencies as they constitute the main medium of information exchange between cryptocurrency market participants. The first Bitcoin transaction was arranged on a forum[1] followed by the first-ever bitcoin related tweet[2] as early as January 11th, 2009. Social media are also an effective way for companies to reach out to potential investors without intermediaries for raising funds through initial coin offerings (ICOs). To put the importance of social media and 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, & Savickas, 2020).

Social media have 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 (Bartov, Faurel, & Mohanram, 2018; Broadstock & Zhang, 2019; Chen, De, Hu, & Hwang, 2014; Gu & Kurov, 2020). Yet, the link between Twitter and cryptocurrency valuations is still unclear. Borri, Massacci, Rubin, and Ruzzi (2022); Liu and Tsyvinski (2021) finds a positive relationship between the number of tweets and future cumulative cryptocurrency returns in the time-series. In the cross-section, results are mixed. Benedetti and Kostovetsky (2021) find evidence supportive of an overreaction channel consistent with Barber and Odean (2008) and Da, Engelberg, and Gao (2011). In contrast, Borri et al. (2022)

[1] https://techcrunch.com/2016/01/02/why-bitcoin-matters/
[2] https://twitter.com/halfin/status/1110302988?refsrctwsrc5Etfw

finds a negative risk premium for investor attention in the cross-section.

The novelty of our paper is that we collect several different samples of tweets for each cryptocurrency. We consider both tweets written by Twitter users and tweets written by the organization developing the cryptocurrency. We retrieve tweets on the cryptocurrency's ticker (*Ticker*-tweets), tweets sent to the cryptocurrency's official account (*Mention*-tweets), and tweets published by the cryptocurrency's official account (*Official*-tweets). Our motivation is that tweets of different types have different characteristics. For instance, *Ticker*-tweets which are 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. This functionality allows *Ticker*-tweets to reach a wider audience than the other types of tweets.

First, we consider the aggregated set of tweets, i.e., *All*-tweets, for each cryptocurrency and create an abnormal attention measure similarly than Da et al. (2011). We show that abnormal attention is positively related to contemporaneous returns and negatively predicts subsequent returns in the cross-section. Following Barillas and Shanken (2017), we show that adding a long-short factor created based on abnormal attention increases the ability of the model of Liu, Tsyvinski, and Wu (2022) to price cryptocurrencies in the cross-section. However, using bivariate sort regression we show that the Twitter factor is only predicting returns in specific sorts which is not consistent with the desirable characteristics of risk factors. Instead, we argue that our results are more consistent with an overreaction channel (Barber & Odean, 2008; Da et al., 2011).

Using our Twitter samples separately, we show that tweets written by Twitter users predict returns consistently with an overreaction channel. Increases in abnormal attention are linked with higher excess returns and trading volume contemporaneously and lower future expected return and trading volume consistent with an increased buying pressure which reverts in the subsequent periods. 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. Furthermore, we show that our results are not due to the events surrounding the GameStop short-squeeze which also caused large abnormal returns and social media activity for cryptocurrencies.

Given that social media are frequently used 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 (2022) who show that cryptocurrency valuations and ICO success are linked to the quality of their underlying technology. Using the daily number of commits on GitHub, as a proxy of 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 any predictability on future returns, the technological innovation channel is not supported by our findings.

Our paper also contributes to the literature on social media. We show that social media content is heterogeneous. Specifically, we document significant differences in the relationships between Twitter abnormal attention and technological innovations or returns depending on who has written the tweets. Our results highlight the need to carefully select the appropriate social media data that is best suited to the desired application. Furthermore, we document that the posting activity of influential users increases the magnitude of the overreaction effect. We argue that like media coverage (Hillert, Jacobs, & Müller, 2014), influential tweets exacerbate behavioral biases. Therefore, our paper contributes to the nascent literature that explores behavioral biases induced by social media activity. For example, social media users tend to self-expose themselves to information in line with their beliefs (Cookson, Engelberg, & Mullins, 2023) and are influenced by investment returns experienced by users they are following (Bailey, Cao, Kuchler, & Stroebel, 2018; Pedersen, 2022).

Section 2 describes the data used in the paper. Section 3 investigates the drivers of Twitter-based attention. Section 4 discusses the interplay of aggregate attention and cryptocurrency returns. Section 5 breaks down attention into several components using the specificity of Twitter and section 6 concludes the paper.

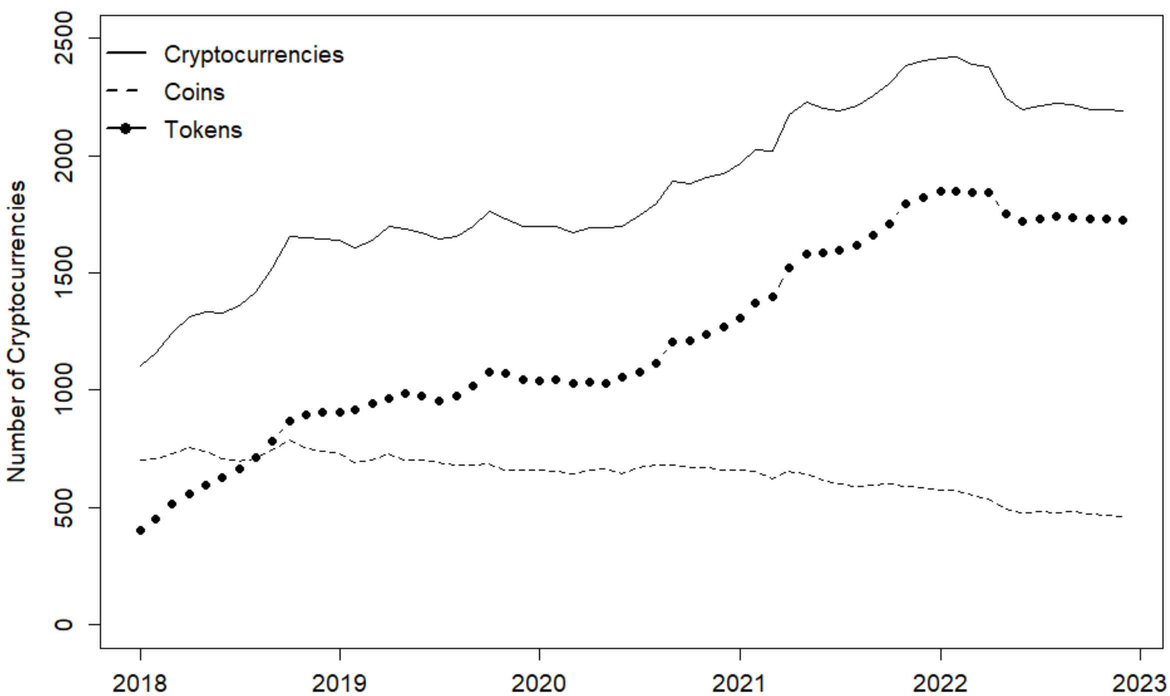
2 Data

2.1 Cryptocurrency data

In this study, we follow the methodology used by Ammann, Burdorf, Liebi, and Stöckl (2022) for collecting and preparing the data of CoinMarketCap. We use an API to download survivorship bias-free data from CoinMarketCap. Our dataset is at a daily frequency in the period from 2018 to 2022. In this paper, we restrict our sample to assets having a market capitalization greater than zero and non-missing volume data. We exclude some cryptocurrencies with erroneously reported data and drop stablecoins which are cryptocurrencies whose value is pegged to other assets such as USD or gold. Our final dataset includes both coins and tokens, which are both types of cryptocurrencies. The former is a means of transaction or store of value like Bitcoin or Dogecoin, while the latter can be exchanged for services on the platform that issued them. For instance, the Edgeless token can be used to play in an online casino. In Figure 1, you can see the average monthly number of cryptocurrencies meeting our criteria.

Figure 1: Development of the Number of Cryptocurrencies over Time

This figure shows the evolution of the number of cryptocurrencies, coins and tokens through time at a monthly frequency.



2.2 Twitter data

Given the download limits imposed by Twitter, we limit our sample to the 161 biggest cryptocurrencies as of the end of the year 2017. We choose this year because it has a large number of cryptocurrencies meeting our criteria. Our sample of 161 assets does not contain cryptocurrencies that have changed name[3] during our sample period or those that do not have a Twitter account. For each of the selected coins, we collect tweets written by the organization developing the cryptocurrency and tweets written by the rest of users separately. We denote the tweets published by those organizations as *Official*-tweets. Collecting tweets generated by users is more complicated, as users have several possibilities to signal that their tweet is about a specific cryptocurrency. Namely, users can use the name, ticker (*Ticker*-tweets) or tag the official account of the cryptocurrency (*Mention*-tweets). As names can be ambiguous, we choose to collect only *Ticker*-tweets and *Mention*-tweets which differ in terms of characteristics and usage. *Mention*-tweets can be viewed as public messages, as the account that is tagged by this type of tweet receives a notification. *Ticker*-tweets have the particularity of being clickable which allows a reader to consult the latest tweets about the specific financial asset being clicked. *Ticker*-tweets have therefore a potential higher reach than the other types of tweets as they can be easily seen by users not following the tweet's author thanks to the clickable feature. Given that users can tweet simultaneously on several cryptocurrencies and can mix the types of tweets, the Twitter samples have some overlap and are not mutually exclusive. Information about the sample size and the characteristics of each samples are summarized in Table 1.

[3] 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 return between the two samples is not statistically significant.

Table 1: Description of Twitter datasets

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 into each sample.

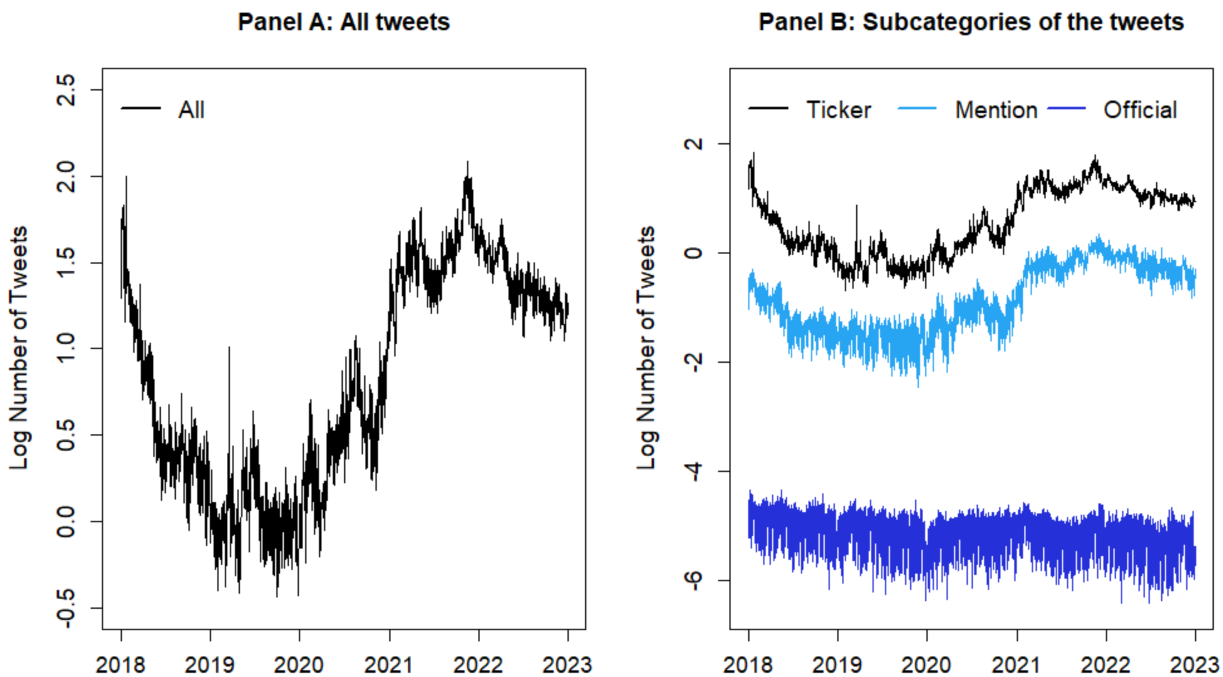
Sample Name	Number of Cryptocurrencies	Of which are dead	Description
Ticker	161	44	Any tweet that contains the ticker of the cryptocurrency
Mention	161	44	Any tweet that contains/mentions the official account of the cryptocurrency
Official	150	35	Tweets published by the official account
All	161	44	The three samples above aggregated

For all samples except sample *Official*, we only download the 100 most relevant tweets[4] per day over the period 2018 to 2022. Sample *Official* includes every tweet posted by the cryptocurrency's official account during the sample period. In all of our samples, the number of Twitter posts published per day and per cryptocurrency is not capped and corresponds to the actual number of tweets published on that date. In Figure 2, you can see the number of tweets posted per sample per day. The number of tweets is scaled by the number of actively traded cryptocurrencies in the Twitter samples. We can see that Twitter activity has increased over our sample period. The only exception is the number of *Official*-tweets which has remained stable over time.

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

Figure 2: Number of Average Tweets over Time

This figure displays the evolution of the average number of tweets of cryptocurrencies through 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 of tweets. Please refer to Table 1 to have more information about the different samples of tweets.



Unfortunately as Twitter provides us with tweets based on textual matching of keywords, our samples of tweets are potentially subject to measurement error. As usernames are unique, we don't expect any measurement error for *Mention*-tweets and *Official*-tweets. However, *Ticker*-tweets have some measurement error, as tickers are generally not unique. Fortunately, this bias doesn't threaten the validity of our results as it only causes standard errors to be larger. Additionally, we don't expect this bias to be large for two reasons. The first is that we consider only tweets written in English which limits the number of ticker homonyms in our sample. Secondly, as cryptocurrencies are a popular topic on social media, we expect the ticker samples to be in majority composed of tweets about cryptocurrencies. As a robustness check, we also report the results if ambiguous tickers are removed from our sample. We judge a ticker as ambiguous if a stock present in the CRSP database has a similar ticker during our sample period.

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. As a consequence, we miss *Official*-tweets from cryptocurrencies with deleted Twitter accounts which are likely to be defunct. However, we do not expect the bias to be large as we could retrieve *Official*-tweets from 35 out of the 44 defunct cryptocurrencies included in our Twitter samples. The survivorship bias is also affecting other samples. But, as the other samples are matching tweets based on their textual content and not based on their issuer, we do not expect survivorship bias to be large. Our rationale is that significant news or event tends to prompt numerous users to tweet. Considering Twitter's extensive user base, the probability of users collaborating to remove all of their past tweets is low.

2.3 GitHub data

To better identify the channels through which Twitter attention predicts the cross-section of expected cryptocurrency returns, we also collect data on each cryptocurrency from GitHub which is the leading platform for collaborating on software development. Specifically, we collect the list of all historical contributions (commits) made by developers on all repositories owned by the organization developing the cryptocurrency. As for *Official*-tweets, GitHub data is also subject to some survivorship bias as cryptocurrencies with missing data are likely defunct. However, we could still collect data for a meaningful set of dead cryptocurrencies. Our GitHub data covers 139 cryptocurrencies and has data for 32 out of the 44 defunct cryptocurrencies in our sample.[5]

2.4 Summary statistics

The main independent variable in this paper is abnormal attention which is closely related to the variable used in Da et al. (2011). This variable is defined as

$$Abn\,Attention_{i,t} = Log(NT_t) - Log(E[NT_{t-30}, NT_{t-29}, ..., NT_{t-1}]) \quad (1)$$

Where NT_t is the number of tweets + 1. This variable as the advantage to be less sensitive to large spikes in number of tweets than the log return of the number of tweets as it takes into account the normal level of attention through its rolling mean component. In order to gauge the activity of important social media users, we construct a dummy variable ($Popular\,Tweet_{i,t}$) equals to one when at least one tweet published per cryptocurrency-date has an aggregated number of likes and replies equal or above 100[6][7]. The variable $Popular\,Tweets_{i,t}$ is not bi-

[5] The numbers do not add up to 161 assets, as some cryptocurrencies use other platforms to publish their code.

[6] The number of retweet 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.

[7] This number corresponds to the 97.5^{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 or replies. Using the number of views of each tweet, we find that the average(median) ratio of views-to-likes is

ased by the fact that we have detailed tweet characteristics for a maximum of 100 tweets per cryptocurrency-date, as we retrieve in priority tweets with the higher number of likes, replies or retweets. Sentiment 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 Loughran and McDonald (2011)[8] which is the main lexicon used for quantifying the sentiment of texts in the financial academic literature. Before computing sentiment, the textual data is first tokenized following Renault (2017). In Table 2, we provide a description of all variables used in our paper.

about 115x(55x) in our sample. 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.

[8]Our results do not change if we use the lexicon of Renault (2017) instead.

Table 2: Variables definition

This table contains a description of the variables used in the paper. CFTC refers to the Commitments of Traders reports issued by the CFTC. 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. CP stands for CoinMarketCap's website. Variables are all at daily frequency unless otherwise mentioned.

Variable	Definition	Source
	Panel A: Cryptocurrency Returns and Characteristics	
$Excess\,Return_{i,t}$	Excess return on day t for cryptocurrency i.	CP, KF
$Size_{i,t}$	Logarithmic market capitalization on day t for cryptocurrency i.	CP
$Volume_{i,t}$	Logarithmic trading volume on day t for cryptocurrency i.	CP
	The variables listed below are 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 cryptocurrency excess return on the daily cryptocurrency market excess return.	CP
$Momentum_{i,t}$	Cumulative performance of cryptocurrency i.	CP
$Volatility_{i,t}$	Standard deviation of returns.	CP
$Idio\,Vola_{i,t}$	Standard deviation of the residuals when daily cryptocurrency excess returns are regressed on daily cryptocurrency market excess returns.	CP
$Max\,Ret_{i,t}$	Average of the five highest daily excess return.	CP
$Volume\,Vola_{i,t}$	Standard deviation of Log transformed trading volume.	CP
$Illiquidity_{i,t}$	Ratio of illiquidity, see Amihud (2002). $Amihud = \frac{1}{T}\Sigma_t^T \frac{\lvert r_{i,t}\rvert}{Volume_{i,t}}$	CP
$Skewness_{i,t}$	Skewness of cryptocurrency daily excess returns.	CP
$Kurtosis_{i,t}$	Kurtosis of cryptocurrency daily excess returns.	CP
$Co-skewness_{i,t}$	The Co-Skewness of cryptocurrency daily excess returns with daily cryptocurrency market excess return. $Coskew = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^2]}{\sigma_{r_i}\sigma_{r_m}^2}$	CP
$Co-kurtosis_{i,t}$	The Co-Kurtosis of cryptocurrency daily excess returns with daily cryptocurrency market excess return. $Cokurt = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^3]}{\sigma_{r_i}\sigma_{r_m}^3}$	CP
$VaR_{i,t}$	The 5% percentile of daily cryptocurrency excess returns.	CP
	Panel B: Attention and Sentiment Measures	
$Abn\,Attention_{i,t}^{All}$	Difference between the logarithm of the total number of tweets at time t and the logarithm of a rolling mean of the number of tweets with a window of 30 days. Please consult equation 1 to see how the variable is defined.	Twitter
$\Delta Attention_{i,t}^{All}$	Difference between the logarithm of the total number of tweets at time t and the logarithm of the number of tweets at time $t-1$.	Twitter
$Sentiment_{i,t}$	Sentiment of the tweets published at date t on cryptocurrency i. We use the Loughran and McDonald (2011) lexicon to compute the sentiment of tweets. Sentiments of individual tweets are then averaged to get a sentiment score at a daily frequency.	Twitter, Author Homepage
$Popular\,Tweet_{i,t}$	Dummy variable equals to one if at least one tweet from any sample published on a cryptocurrency-date has a aggregated number of likes and replies equal or higher than 100.	Twitter
	Panel C: Technology Improvement Measures	
$Commit_{i,t}$	Logarithmic return of the number of commits + 1 published per day on GitHub for each repositories of the organization developing the respective cryptocurrency.	GitHub

We follow Liu, Tsyvinski, and Wu (2022) to construct the market, size and momentum factors. The returns are trimmed at the 2% level and the remaining variables are winsorized at the 1% level. Returns are trimmed to alleviate the effects of suspicious data on the results (Ammann et al., 2022). In Table 3, you can see some descriptive statistics for the variables used in this paper.

Table 3: Summary Statistics

This table contains the summary statistics of the variables defined in Table 2. All variables are winsorized at the 1% level, except for $Excess\,Returns_t$ which is trimmed at the 2% level. The variables are expressed in decimal points. Variables are all at daily frequency.

	$Mean$	25%	$Median$	75%	$StdDev$
Panel A: Cryptocurrency Returns and Characteristics					
$Excess\,Return_{i,t}$	−0.0001	−0.0368	−0.002	0.0321	0.0774
$Beta_{i,t}$	0.9336	0.7553	0.9635	1.1448	0.33
$Size_{i,t}$	17.3854	15.7248	17.0642	18.7831	2.5236
$Momentum_{i,t}$	0.0063	−0.3951	−0.1402	0.1899	0.6637
$Volatility_{i,t}$	0.0731	0.0517	0.067	0.0884	0.0295
$Idio\,Vola_{i,t}$	0.0573	0.0345	0.05	0.0734	0.0318
$Max\,Ret_{i,t}$	0.1469	0.0952	0.131	0.1836	0.0692
$Volume_{i,t}$	−0.0033	−0.308	−0.023	0.2561	0.7298
$Volume\,Vola_{i,t}$	0.792	0.4643	0.694	0.9762	0.4927
$Illiquidity_{i,t}$	0.0014	0	0	0	0.0122
$Skewness_{i,t}$	0.348	−0.1427	0.298	0.7671	0.7796
$Kurtosis_{i,t}$	5.1734	3.4613	4.3362	5.8949	2.6713
$Co\text{-}Skewness_{i,t}$	−0.2512	−0.5025	−0.2271	0.0158	0.4242
$Co\text{-}Kurtosis_{i,t}$	2.737	1.7124	2.5086	3.3719	1.6734
$VaR_{i,t}$	−0.1066	−0.1283	−0.0981	−0.0753	0.0441
Panel B: Attention and Sentiment Measures					
$Abn\,Attention^{All}_{i,t}$	−0.2164	−0.5717	−0.1573	0.1537	0.6873
$Abn\,Attention^{Mention}_{i,t}$	−0.2812	−0.6994	−0.1054	0.0225	0.7934
$Abn\,Attention^{Official}_{i,t}$	−0.1083	−0.3365	−0.0328	0	0.4657
$Abn\,Attention^{Ticker}_{i,t}$	−0.207	−0.5448	−0.1384	0.1422	0.6773
$Sentiment^{All}_{i,t}$	0.0778	0	0.0526	0.1824	0.1878
$Popular\,Tweet_{i,t}$	0.165	0	0	0	0.3712
Panel C: Technology Improvement Measures					
$Commit_{i,t}$	−0.0004	−0.2231	0	0.1823	0.8578

3 Determinants of Twitter-Based Investor Attention

We first investigate what are the drivers of Twitter-based investor attention. In Table 4, we run a panel regression of contemporaneous and future changes in attention for each Twitter sample on changes in attention respectively reported in Panel A and B. 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 2. In Panel A of Table 4, we see that all attention variables are all positively related to each other at the 1% significance level. Daily excess returns are also significantly linked with contemporaneous change in attention for *Mention*-tweets and *Ticker*-tweets. This observation is consistent with the results of Da et al. (2011) which also find that investor attention is predictable by returns. However, the absence of relationship between returns and *Official*-tweets suggests that cryptocurrencies in average do not time their announcements to manipulate their prices. $Popular\,Tweet_{i,t}$ predicts also positively contemporaneous changes in attention. We interpret this finding as evidence that impacting tweets induce more users to tweet. In Panel B, we can see that future changes in attention of each samples are inversely related with their respective lagged change in attention at the 1% significance level. Attention tends to revert to a normal level following spikes or drops in attention. This observation is further confirmed by the negative sign of $Popular\,Tweet_{i,t}$ across the three models. Future attention is significantly linked with previous day excess returns for *Ticker*-tweets and *Mention*-tweets but not for *Official*-tweets. The link between lagged returns and Twitter attention is the strongest for *Ticker*-tweets both in terms of statistical and economical significance. Intuitively this observation makes sense given that *Ticker*-tweets are mainly used for talking about trading and investments. Additionally, we can see that changes in the number of *Official*-tweets positively predict future attention for the other samples. This evidence is consistent with the view that social media content is heterogeneous. *Official*-tweets seems to represent news to which Twitter users are reacting by posting *Ticker*-tweets and *Mention*-tweets.

Table 4: Determinants of refined Twitter-Based Attention

The dependent variable is change in attention using several Twitter samples. The regression spans the time period 2018 to 2022 for a sample of 150 cryptocurrencies. Control variables are defined in Panel A of Table 2. 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% respectively.

	Panel A: Daily Contemporaneous Relationship		
	(1)	(2)	(3)
	$\Delta Attention_t^{Mention}$	$\Delta Attention_t^{Official}$	$\Delta Attention_t^{Ticker}$
$\Delta Attention_{i,t}^{Mention}$		23.37***	10.04***
		(26.12)	(16.49)
$\Delta Attention_{i,t}^{Official}$	47.76***		1.63***
	(21.44)		(3.17)
$\Delta Attention_{i,t}^{Ticker}$	20.49***	1.63***	
	(14.95)	(3.17)	
$Excess\ Return_{i,t}$	34.21***	−2.14	70.19***
	(9.04)	(−1.08)	(12.01)
$Popular\ Tweet_{i,t}$	21.42***	2.61***	19.22***
	(13.29)	(3.12)	(15.57)
Controls	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	238483	238483	238483
	Panel B: Daily Predictive Relationship		
	(1)	(2)	(3)
	$\Delta Attention_{t+1}^{Mention}$	$\Delta Attention_{t+1}^{Official}$	$\Delta Attention_{t+1}^{Ticker}$
$\Delta Attention_{i,t}^{Mention}$	−36.79***	−1.83***	−0.13
	(−81.72)	(−10.36)	(−0.79)
$\Delta Attention_{i,t}^{Official}$	10.1***	−41.96***	1.22***
	(14.29)	(−80.34)	(4.89)
$\Delta Attention_{i,t}^{Ticker}$	2.37***	0.44*	−32.61***
	(5.95)	(1.9)	(−66.28)
$Excess\ Return_{i,t}$	23.59***	2.19	35.3***
	(7.26)	(1.31)	(9.48)
$Popular\ Tweet_{i,t}$	−5.11***	−3.85***	−3.57***
	(−6.43)	(−6.84)	(−6.98)
Controls	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	238371	238371	238371

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

We now study the link between abnormal attention (sample *All*) and the cross-section of cryptocurrency expected returns. Any tweet being in at least one of the three samples of tweets is included in sample *All*. Therefore, for cryptocurrencies with deleted official account *All*-tweets only include *Mention*-tweets and *Ticker*-tweets.

4.1 Portfolio Sorts

Table 5 reports the results of portfolio sorts regressions. Every day, we sort cryptocurrencies into terciles which are then equally-weighted to create portfolios. We report the performance of the portfolios in time t and $t+1$ in Panel A and Panel B respectively. In Panel C, we form portfolios every week and report their performances at time $t+1$. We can see that the long-short portfolio that is long cryptocurrencies with a high change in the number of tweets and short cryptocurrencies with a low change in the number of tweets is not priced by the model of Liu, Tsyvinski, and Wu (2022) in Panel A, B and C. The long-short portfolio has a daily alpha of -0.08% per day in Panel B. When annualized, the alpha of panel B is about -29% which is economically large. However, the alphas of the long-short portfolios created do not have all the same sign. Change in attention doesn't uniformly predict contemporaneous and future returns which suggest that change in attention is more likely to be an anomaly than a new risk factor for cryptocurrencies.

From an investment point of view, it is unlikely that our attention portfolios can be easily implemented in practices. Short-selling is hard to implement for most cryptocurrencies. For instance, on the CME, trading futures contracts is only possible for Ethereum and Bitcoin. Furthermore, the alphas shown in Table 5 do not account for transaction costs which are likely to be large given the frequent re-balancing required by the trading strategy to work.

Table 5: Univariate Portfolio Sorts

Cryptocurrencies are sorted into terciles based on $Abn\,Attention_t^{All}$. The coefficients reported in the table are expressed in percentage points. Standard errors are adjusted for serial-correlation and have a bandwidth of 3. The t-statistics are reported in parenthesis below the coefficients. ***, **, * indicate respectively statistical significance at the 1%, 5% and 10% respectively.

	Panel A: Equal-Weighted Contemporaneous Relationship			
	$Abn\,Attention_t^{All}$	$Excess\,Return_t$	$One\text{-}Factor\,Alpha_t$	$Liu\,et\,Al.\,Alpha_t$
$Q1$	−88.55***	−0.48***	−0.46***	−0.5***
	(−135.32)	(−5.03)	(−15.75)	(−17.71)
$Q2$	−18.83***	−0.21**	−0.19***	−0.22***
	(−39.19)	(−2.11)	(−7.07)	(−8.32)
$Q3$	43.51***	0.7***	0.72***	0.68***
	(82.33)	(6.95)	(18.44)	(18.1)
$Q3 - Q1$	132.06***	1.18***	1.18***	1.18***
	(233.72)	(31.13)	(28.53)	(28.87)
	Panel B: Daily Equal-Weighted Predictive Relationship			
	$Abn\,Attention_{t+1}^{All}$	$Excess\,Return_{t+1}$	$One\text{-}Factor\,Alpha_{t+1}$	$Liu\,et\,Al.\,Alpha_{t+1}$
$Q1$	−88.54***	0.02	0.04	0
	(−135.26)	(0.17)	(1.17)	(−0.14)
$Q2$	−18.83***	0.06	0.08***	0.05*
	(−39.16)	(0.57)	(2.81)	(1.91)
$Q3$	43.51***	−0.08	−0.05*	−0.09***
	(82.28)	(−0.79)	(−1.87)	(−3.13)
$Q3 - Q1$	132.04***	−0.09***	−0.09***	−0.08**
	(233.66)	(−2.64)	(−2.76)	(−2.43)
	Panel C: Weekly Equal-Weighted Predictive Relationship			
	$Abn\,Attention_{t+1}^{All}$	$Excess\,Return_{t+1}$	$One\text{-}Factor\,Alpha_{t+1}$	$Liu\,et\,Al.\,Alpha_{t+1}$
$Q1$	−86.6***	−0.68	−0.41	−0.81**
	(−57.81)	(−0.96)	(−1.61)	(−2.57)
$Q2$	−17.46***	−0.22	0.06	−0.33
	(−15.13)	(−0.3)	(0.22)	(−1.1)
$Q3$	44.34***	−0.22	0.04	−0.23
	(31.56)	(−0.33)	(0.15)	(−0.71)
$Q3 - Q1$	130.93***	0.46**	0.45**	0.58***
	(92.16)	(2.07)	(2.37)	(2.91)

We now investigate if the effects of change of attention are the same across subgroups of cryptocurrencies. In Table 6, we report the results of dependent double portfolio sort regressions. Each period, all cryptocurrencies are first sorted on $Beta_t$, $Size_t$ or $Momentum_t$. Then, cryptocurrencies are sorted into terciles within each group based on $Abn\,Attention_t^{All}$. The 2x3 sorts are then equally-weighted to form portfolios. In Panel A, we report the performance at time t of the portfolios formed at time t using daily data. In Panel B, we report the daily performance of the portfolios at time $t+1$ instead. In Panel C, we use weekly data to form portfolios at time t and report their performances at time $t+1$. We can see that forming long-short portfolios based on $Abn\,Attention_t^{All}$ generates strongly positive alphas at time t in all sorts. In Panel B, the sign of alphas flip and alphas are statistically significant only in the high beta, high size and in the high momentum group. In Panel C, the alpha is statistically significant in the high and low momentum group which is consistent with the results of Hillert et al. (2014) who finds that momentum effect can be exacerbated for stocks with relatively larger media coverage. Since the alphas are not pervasive across the different sorts, it provides additional evidence that our long-short portfolio formed on $Abn\,Attention_t^{All}$ is not a new risk factor. To get a better understanding of why our Twitter-based attention variable can predict returns in some sorts, we run panel regressions and other tests in the next sections of the paper.

Table 6: Bivariate Portfolio Sorts

Each period cryptocurrencies are sorted dependently in six portfolios based on the following variables: $Abn\,Attention_t^{All}$, $Beta_t$, $Size_t$ and $Momentum_t$. The cryptocurrencies are first sorted based on $Beta_t$, $Size_t$ or $Momentum_t$ in two groups, then assets are sorted within each group based on $Abn\,Attention_t^{All}$ in three groups. The portfolios are equally-weighted. The table reports returns and alphas in percentage points. Standard errors are adjusted for serial-correlation and have a bandwidth of 3. The t-statistics are reported in parenthesis below the coefficients. ***, **, * indicate respectively statistical significance at the 1%, 5% and 10% respectively.

	$Low\,Beta_t$	$High\,Beta_t$	$Low\,Size_t$	$High\,Size_t$	$Low\,Momentum_t$	$High\,Momentum_t$
Panel A: Contemporaneous Relationship						
$Q1\,Abn\,Attention_t^{All}$	−0.42***	−0.56***	−0.4***	−0.61***	−0.85***	−0.08
	(−4.95)	(−4.8)	(−4.23)	(−5.76)	(−8.7)	(−0.79)
$Q2\,Abn\,Attention_t^{All}$	−0.21**	−0.21*	−0.16*	−0.2*	−0.62***	0.2**
	(−2.5)	(−1.77)	(−1.7)	(−1.87)	(−6.14)	(1.97)
$Q3\,Abn\,Attention_t^{All}$	0.51***	0.85***	0.3***	1.07***	−0.33***	1.64***
	(6.08)	(6.72)	(3.07)	(9.67)	(−3.24)	(15.32)
$Q3 - Q1$	0.92***	1.41***	0.7***	1.67***	0.53***	1.72***
	(17.68)	(30.47)	(12.41)	(38.82)	(11.23)	(32.02)
Liu et Al. Alpha	0.93***	1.41***	0.69***	1.67***	0.53***	1.73***
	(17.72)	(27.48)	(12.08)	(33.57)	(12.12)	(28.54)
Panel B: Daily Predictive Relationship						
	$Low\,Beta_t$	$High\,Beta_t$	$Low\,Size_t$	$High\,Size_t$	$Low\,Momentum_t$	$High\,Momentum_t$
$Q1\,Abn\,Attention_t^{All}$	−0.03	0.07	0.05	−0.01	0.16	−0.13
	(−0.35)	(0.66)	(0.5)	(−0.12)	(1.59)	(−1.3)
$Q2\,Abn\,Attention_t^{All}$	0.03	0.04	0.08	0.02	0.16	−0.1
	(0.32)	(0.34)	(0.79)	(0.16)	(1.56)	(−0.97)
$Q3\,Abn\,Attention_t^{All}$	−0.08	−0.07	0.02	−0.16	0.12	−0.25**
	(−0.91)	(−0.66)	(0.25)	(−1.54)	(1.15)	(−2.42)
$Q3 - Q1$	−0.05	−0.15***	−0.02	−0.15***	−0.04	−0.11**
	(−1.01)	(−3.35)	(−0.42)	(−3.99)	(−0.89)	(−2.45)
Liu et Al. Alpha	−0.04	−0.14***	−0.03	−0.14***	−0.04	−0.1**
	(−0.92)	(−3.17)	(−0.56)	(−3.83)	(−0.8)	(−2.32)
Panel C: Weekly Predictive Relationship						
	$Low\,Beta_t$	$High\,Beta_t$	$Low\,Size_t$	$High\,Size_t$	$Low\,Momentum_t$	$High\,Momentum_t$
$Q1\,Abn\,Attention_t^{All}$	−1	−0.4	−0.58	−0.75	−0.39	−0.99
	(−1.49)	(−0.49)	(−0.84)	(−0.95)	(−0.52)	(−1.36)
$Q2\,Abn\,Attention_t^{All}$	−0.45	−0.04	−0.31	0.12	0.19	−0.69
	(−0.69)	(−0.05)	(−0.45)	(0.16)	(0.26)	(−0.92)
$Q3\,Abn\,Attention_t^{All}$	−0.09	−0.32	−0.24	−0.5	0.17	−0.64
	(−0.14)	(−0.41)	(−0.35)	(−0.68)	(0.24)	(−0.9)
$Q3 - Q1$	0.91***	0.08	0.35	0.25	0.56*	0.36
	(3.21)	(0.26)	(1.18)	(0.9)	(1.92)	(1.3)
Liu et Al. Alpha	0.97***	0.19	0.34	0.36	0.62**	0.48*
	(3.97)	(0.63)	(1.21)	(1.4)	(2.38)	(1.84)

In Table 7, we test whether adding our Twitter-based long-short portfolio to the model of Liu, Tsyvinski, and Wu (2022) improves is ability to price test assets or not. Each week, we form equally-weighted long-short portfolios based on *Beta*, *Size* and *Momentum* using data available at time t. We then record the performance of those long-short portfolios at time

$t+1$. We report the mean returns, alphas based on Liu, Tsyvinski, and Wu (2022) model and the alphas based on Liu, Tsyvinski, and Wu (2022) augmented by our Twitter-based long-short portfolio. We find that the addition of our Twitter-based portfolio is helpful in decreasing the alphas of the test assets. However, the difference is not large suggesting that our Twitter-based portfolio is not an helpful risk factor.

Table 7: Pricing Test Assets

Each week cryptocurrencies are sorted in terciles based on the following variables: $Beta_t$, $Size_t$ and $Momentum_t$. The sorts are equally-weighted to form portfolios. The table reports returns and alphas in percentage points. Standard errors are adjusted for serial-correlation and have a bandwidth of 6. The t-statistics are reported in parenthesis below the coefficients. ***, **, * indicate respectively statistical significance at the 1%, 5% and 10% respectively.

	(1) *Beta*	(2) *Size*	(3) *Momentum*
$Q1$	-3.22	-9.33^{**}	-24.07^{***}
	(-0.81)	(-2.06)	(-5.05)
$Q2$	-3.34	-0.05	-5.72
	(-0.67)	(-0.01)	(-1.13)
$Q3$	-0.95	2.28	22.12^{***}
	(-0.16)	(0.44)	(4.27)
$Q3 - Q1$	2.27	11.61^{***}	46.19^{***}
	(0.8)	(5.07)	(17.79)
Liu et Al. Alpha	2.79	13.57^{***}	47.19^{***}
	(1.41)	(5.99)	(18.98)
Liu et Al. + Attention Alpha	2.53	2.5	40.16^{***}
	(0.99)	(0.85)	(12.07)

4.2 Multivariate Analysis

In Table 8, we present the results of panel regressions of excess returns on abnormal attention and various controls as defined in Table 2. The panel regressions allow us to control for additional variables compared to the portfolios sort. The first model investigates the link between Twitter attention and returns contemporaneously. The other models replace contemporaneous returns by future returns using different time horizons. As expected, we note that the results from the sample containing all the cryptocurrencies in Panel A are similar to the results of Panel B where cryptocurrencies with ambiguous tickers have been removed. We can see that the regression coefficient of $Abn\,Attention_t^{All}$ is statistically significant in model (1) at the 1% significance level and in models (2) and (6) at the 5% significance level. The regression coefficient of $Abn\,Attention_t^{All}$ is positive in model (1) and (6) and then negative in model (2). An increase in Twitter-based attention predicts higher contemporaneous returns and negatively the subsequent returns which is consistent with an overreaction channel (Barber & Odean, 2008; Da et al., 2011). However, we need to be cautious when interpreting the results of Table 8 as Twitter data may predict future expected returns through several non-exclusive channels. For instance, *Official*-tweets are included in sample *All* although they differ conceptually from other types of tweets as shown in Table 4. It is likely that *Official*-tweets which can be seen as corporate announcements predict cryptocurrency returns differently than other types of tweets which can be posted by anyone.

Table 8: Multivariate Analysis: Contemporaneous Returns, Future Returns and Attention

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the time period 2018 to 2022 for a sample of 161 cryptocurrencies. Control variables are defined in Panel A of Table 2. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parenthesis below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% respectively.

	Panel A: All Cryptocurrencies					
	(1) $Excess\ Return_{i,t}$	(2) $Excess\ Return_{i,t+1}$	(3) $Excess\ Return_{i,t+2}$	(4) $Excess\ Return_{i,t+3}$	(5) $Excess\ Return_{i,t+4}$	(6) $Excess\ Return_{i,t+5}$
$Abn\,Attention_{i,t}^{All}$	0.67***	−0.08**	−0.03	0.01	0.04	0.04**
	(16.28)	(−2.47)	(−1.35)	(0.24)	(1.51)	(2)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249743	249139	248614	248384	248157	247978
	Panel B: Filtered Ticker					
	(1) $Excess\ Return_{i,t}$	(2) $Excess\ Return_{i,t+1}$	(3) $Excess\ Return_{i,t+2}$	(4) $Excess\ Return_{i,t+3}$	(5) $Excess\ Return_{i,t+4}$	(6) $Excess\ Return_{i,t+5}$
$Abn\,Attention_{i,t}^{All}$	0.69***	−0.09**	−0.02	0.02	0.04	0.04*
	(14.39)	(−2.29)	(−0.76)	(0.71)	(1.48)	(1.74)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	161354	160978	160657	160513	160371	160249

4.3 Robustness Checks

To show the robustness of our results, we estimate panel regression models with slight modifications compared to Table 8. The robustness checks are reported in Table 9. In column one, we consider raw attention instead of abnormal attention which is defined as:

$$Raw\,Attention_{i,t} = Log(1 + Number\,Of\,Tweets_{i,t}) \tag{2}$$

In the other columns, we use abnormal attention as our main independent variable as in Table 8. In models (2) and (3), the sample is using only cryptocurrencies classified as coins or as tokens respectively. In models (4) and (5), the sample only contains observations anterior or posterior respectively to the GameStop short-squeeze. In model (6) and (7), we restrict our sample to assets with a market capitalization respectively above or below the median market capitalization. In all model specifications of Panel A, the regression coefficients of Twitter attention are 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 reversal is statistically significant for large coins but not for small coins. This is consistent with a limits-to-arbitrage explanation, as larger cryptocurrencies are easier to short-sell than smaller assets. We also check if the results are impacted by the rally of meme stocks led by R/WallStreetBets in 2021 which also affected some cryptocurrencies [9]. We obtain similar results in the sample ending(starting) before(after) the first three months of 2021. This finding provides evidence that our results are not driven by this particular event. Then, we can see that the predictability of attention is slightly stronger for tokens than for coins in Panel B.

[9] https://www.coindesk.com/markets/2021/02/16/what-really-happened-when-robinhood-suspended-gamestop-trading/

However, the difference is not big. Lastly, we can see that raw attention does not predict any reversals in t+1. This is not surprising as raw attention is a noisy measure of attention shocks.

Table 9: Robustness Checks

The dependent variable is daily excess returns ($Excess\,Return_t$). The regression spans the time period 2018 to 2022 for a sample of 161 cryptocurrencies. Control variables are defined in Panel A of Table 2. The first model (1) uses raw attention instead of abnormal change as 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 anterior or posterior to the GameStop short-squeeze. Finally, models (6) and (7) restricts the sample to assets with a market cap below or above the median respectively. 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.

Panel A: Contemporaneous Relationship
Dependent variable: $Excess\,Return_{i,t}$

	(1) *Raw Attention*	(2) *Coins Only*	(3) *Tokens Only*	(4) *Before GameStop*	(5) *After GameStop*	(6) *Large Only*	(7) *Small Only*
$Abn\,Attention^{All}_{i,t}$	0.16***	0.67***	0.66***	0.75***	0.52***	1.09***	0.37***
	(7.28)	(12.91)	(10.48)	(15.55)	(9.48)	(19.78)	(9.39)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249743	142914	106829	161644	77826	126354	122576

Panel B: Predictive Relationship $t+1$
Dependent variable: $Excess\,Return_{i,t+1}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Abn\,Attention^{All}_{i,t}$	0.01	−0.07	−0.11**	−0.07*	−0.11**	−0.18***	−0.03
	(0.6)	(−1.6)	(−2.4)	(−1.88)	(−2.42)	(−3.86)	(−0.73)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249139	142600	106539	161387	77504	126314	122013

5 Additional Empirical Results

5.1 Twitter-Based Sentiment and the Cross-Section of Cryptocurrency Returns

To get a better understanding on why $Abn\,Attention^{All}_{i,t}$ predicts the cross-section of expected returns, we test whether the results change if we account for the sentiment of tweets. In

Table 10, we report the results of a panel regression using both attention and sentiment as independent variables. We also include an interaction term to test whether the potential over-reaction effect is stronger when an attention grabbing tweet is posted. The regression coefficients of $Abn\,Attention_{i,t}^{All}$ are robust to the inclusion of the two additional independent variables. In Panel A, $Sentiment_{i,t}^{All}$ is positively linked with contemporaneous and next day returns at the 1% significance level which is in line with previous literature on Twitter sentiment (Gu & Kurov, 2020; Jiang, Liu, Roch, & Zhou, 2023). We can see that the regression coefficients of the interaction term is reinforcing the effect of abnormal attention on returns. If a popular tweet is posted, the overreaction effect is exacerbated. This finding echoes with Hillert et al. (2014) which 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. Overall, evidence contained in Table 10 suggests that Twitter attention is primary affecting the cross-section of expected returns through an overreaction channel.

Table 10: Multivariate Analysis: Contemporaneous Returns, Future Returns, Attention, and Sentiment

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the time period 2018 to 2022 for a sample of 161 cryptocurrencies. Control variables are defined in Panel A of Table 2. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parenthesis below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% respectively.

Panel A: All Cryptocurrencies

	(1) Excess $Return_{i,t}$	(2) Excess $Return_{i,t+1}$	(3) Excess $Return_{i,t+2}$	(4) Excess $Return_{i,t+3}$	(5) Excess $Return_{i,t+4}$	(6) Excess $Return_{i,t+5}$
$Abn\,Attention^{All}_{i,t}$	0.54***	−0.06*	−0.01	0.02	0.04*	0.05**
	(15.81)	(−1.74)	(−0.57)	(0.67)	(1.65)	(2.41)
$Sentiment^{All}_{i,t}$	0.2***	0.31***	0.05	0.05	0.13*	0.08
	(2.62)	(4.17)	(0.7)	(0.63)	(1.73)	(1.13)
$Abn\,Attention^{All}_{i,t}X$	1.06***	−0.22***	−0.15**	−0.09	−0.06	−0.11*
$Popular\,Tweet_{i,t}$	(8.28)	(−2.68)	(−2.39)	(−1.49)	(−0.98)	(−1.74)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	249743	249139	248614	248384	248157	247978

Panel B: Filtered Ticker

	(1) Excess $Return_{i,t}$	(2) Excess $Return_{i,t+1}$	(3) Excess $Return_{i,t+2}$	(4) Excess $Return_{i,t+3}$	(5) Excess $Return_{i,t+4}$	(6) Excess $Return_{i,t+5}$
$Abn\,Attention^{All}_{i,t}$	0.55***	−0.07*	−0.01	0.03	0.04	0.06**
	(14.35)	(−1.78)	(−0.33)	(1.06)	(1.38)	(2.02)
$Sentiment^{All}_{i,t}$	0.23**	0.26***	0.09	0.03	0.11	0.11
	(2.36)	(3.1)	(1.09)	(0.29)	(1.17)	(1.34)
$Abn\,Attention^{All}_{i,t}X$	1.12***	−0.19*	−0.1	−0.1	−0.01	−0.13
$Popular\,Tweet_{i,t}$	(7.18)	(−1.92)	(−1.35)	(−1.46)	(−0.15)	(−1.59)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	161354	160978	160657	160513	160371	160249

5.2 Refinements of Investor Attention

As each category of tweets differs in terms of types of authors, functionalities or reach, we have good reasons to 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 to trigger a notification for the recipient of the mention. *Ticker*-tweets upon being clicked display the more recent *Ticker*-tweets about the corresponding financial asset. We now study if the qualitative differences between our Twitter samples translate into different relationships in the cross-section of cryptocurrency expected returns. In Table 11, we estimate panel regressions of excess return on the Twitter samples while controlling for various variables. We can see that *Ticker*-tweets and *Mention*-tweets are both consistent with an overreaction channel. Those two variables predict positively contemporaneous and negatively returns in at least one subsequent period. In contrast, $Abn\,Attention_{i,t}^{Official}$ doesn't predict returns in any of the panels. The results of Table 11 illustrate that the overreaction effect comes mainly from user-generated content and not from the announcements made by cryptocurrencies.

Table 11: Multivariate Regressions with Refinements of Investor Attention

The dependent variable is daily returns ($Excess\,Return_t$). The regression spans the time period 2018 to 2022 for a sample of 150 cryptocurrencies. Control variables are defined in Panel A of Table 2. 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.

Panel A: All Cryptocurrencies

	(1) *Excess* $Return_{i,t}$	(2) *Excess* $Return_{i,t+1}$	(3) *Excess* $Return_{i,t+2}$	(4) *Excess* $Return_{i,t+3}$	(5) *Excess* $Return_{i,t+4}$	(6) *Excess* $Return_{i,t+5}$
$Abn\,Attention_{i,t}^{Mention}$	0.19***	−0.05**	−0.03	−0.02	0	−0.01
	(9.33)	(−2.45)	(−1.48)	(−1.22)	(0.12)	(−0.6)
$Abn\,Attention_{i,t}^{Official}$	0.05	0.02	0	0.03	0	0.01
	(1.55)	(0.65)	(−0.16)	(1.03)	(−0.1)	(0.24)
$Abn\,Attention_{i,t}^{Ticker}$	0.7***	−0.07**	−0.02	0	0.01	0.04*
	(14.24)	(−2.11)	(−0.94)	(0.07)	(0.52)	(1.91)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	238279	237755	237294	237080	236877	236710

Panel B: Filtered Ticker

	(1) *Excess* $Return_{i,t}$	(2) *Excess* $Return_{i,t+1}$	(3) *Excess* $Return_{i,t+2}$	(4) *Excess* $Return_{i,t+3}$	(5) *Excess* $Return_{i,t+4}$	(6) *Excess* $Return_{i,t+5}$
$Abn\,Attention_{i,t}^{Mention}$	0.19***	−0.08***	−0.04*	−0.01	0	0
	(7.71)	(−3.1)	(−1.77)	(−0.36)	(−0.13)	(−0.09)
$Abn\,Attention_{i,t}^{Official}$	0.05	0.03	−0.01	−0.01	0.02	0.01
	(1.49)	(0.95)	(−0.38)	(−0.2)	(0.55)	(0.22)
$Abn\,Attention_{i,t}^{Ticker}$	0.72***	−0.06	−0.01	0.01	0.01	0.04
	(12.34)	(−1.33)	(−0.24)	(0.19)	(0.47)	(1.48)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	157713	157370	157068	156921	156789	156675

We now investigate how Twitter-based attention predicts contemporaneous and future change in trading volume in a panel regression setting. Results are reported in Table 12. *Mention*-tweets and *Ticker*-tweets are both strongly linked with contemporaneous and future change in volume up to $t+4$. Both types of tweets are positively linked with contemporaneous volume at the 1% significance level and negatively 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 which decreases over the subsequent days.

Table 12: Multivariate Regressions of volume with Refinements of Investor Attention

The dependent variable is daily log-change in trading volume ($Volume_t$). The regression spans the time period 2018 to 2022 for a sample of 150 cryptocurrencies. Control variables are defined in Panel A of Table 2. 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.

Panel A: All Cryptocurrencies

	(1) $Volume_{i,t}$	(2) $Volume_{i,t+1}$	(3) $Volume_{i,t+2}$	(4) $Volume_{i,t+3}$	(5) $Volume_{i,t+4}$	(6) $Volume_{i,t+5}$
$Abn\ Attention_{i,t}^{Mention}$	1.58***	−1.61***	−0.95***	−0.34**	−0.37**	−0.36**
	(7.09)	(−8.26)	(−5.07)	(−2.19)	(−2.01)	(−2.28)
$Abn\ Attention_{i,t}^{Official}$	0.57*	0.82***	−0.57*	0.6**	0.06	−0.16
	(1.74)	(2.8)	(−1.86)	(2.16)	(0.21)	(−0.54)
$Abn\ Attention_{i,t}^{Ticker}$	8.76***	−4.12***	−2.65***	−0.9***	−0.38*	−0.27
	(12.8)	(−10.51)	(−8.4)	(−4.4)	(−1.7)	(−1.44)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	240178	239629	239140	238904	238691	238514

Panel B: Filtered Ticker

	(1) $Volume_{i,t}$	(2) $Volume_{i,t+1}$	(3) $Volume_{i,t+2}$	(4) $Volume_{i,t+3}$	(5) $Volume_{i,t+4}$	(6) $Volume_{i,t+5}$
$Abn\ Attention_{i,t}^{Mention}$	1.59***	−1.87***	−0.82***	−0.39*	−0.51**	−0.4*
	(5.45)	(−8.12)	(−3.38)	(−1.92)	(−2.1)	(−1.87)
$Abn\ Attention_{i,t}^{Official}$	0.65	1.04***	−0.74**	0.79**	0.1	−0.14
	(1.62)	(2.72)	(−2.11)	(2.16)	(0.29)	(−0.4)
$Abn\ Attention_{i,t}^{Ticker}$	8.91***	−4.54***	−2.5***	−0.83***	−0.57**	−0.05
	(10.42)	(−9.37)	(−6.41)	(−3.49)	(−2.48)	(−0.2)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	158841	158485	158170	158009	157869	157751

Finally, we also test whether Twitter-based attention also predicts expected returns through a technological innovation channel. Twitter content could potentially predicts expected returns through its predictability on technological improvement on the blockchain. This channel is plausible as social media are frequently used by developers to exchange ideas about potential improvements or to signal cybersecurity breaches. In the literature, Cong, Li, and Wang (2021) show theoretically that cryptocurrencies valuations are linked with technological improvements. Empirically, technological innovation is linked negatively with delisting probability (Liu, Sheng, & Wang, 2022) and positively with ICO success (Lyandres, Palazzo, & Rabetti, 2022). To proxy for technological improvements, we use the number of commits published for each date and cryptocurrency on GitHub. We think this proxy is good as a commit captures any significant code changes 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. The results are reported in Table 13. We can see that contemporaneous and future cryptocurrency's technological development is strongly linked with *Official*-tweets. The regression coefficients of *Official*-tweets are statistically significant at the 1% level in models $(1)-(3)$. In contrast, the relationship between abnormal attention based on user-generated content and GitHub activity is weak. Those results clearly highlight that *Ticker*-tweets and *Mention*-tweets predict return through an overreaction channel and not through a technology innovation channel. The significant link between the number of Official-tweets with technological innovation further confirm that tweets posted by cryptocurrencies can be interpreted as news.

Table 13: Multivariate Regressions of commits with Refinements of Investor Attention

The dependent variable is daily log-change in the number of commits ($Commit_t$). The regression spans the time period 2018 to 2022 for a sample of 132 cryptocurrencies. Control variables are defined in Panel A of Table 2. 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)	(5)	(6)
Panel A: All Cryptocurrencies						
	$Commit_{i,t}$	$Commit_{i,t+1}$	$Commit_{i,t+2}$	$Commit_{i,t+3}$	$Commit_{i,t+4}$	$Commit_{i,t+5}$
$Abn\ Attention^{Mention}_{i,t}$	−0.21	−0.48	−0.12	−0.13	−0.26	0.17
	(−0.69)	(−1.13)	(−0.37)	(−0.63)	(−1.27)	(0.69)
$Abn\ Attention^{Official}_{i,t}$	5.56***	−6.61***	−5.82***	−0.72	0.24	0.74
	(5.24)	(−6.41)	(−5.21)	(−1.45)	(0.55)	(1.33)
$Abn\ Attention^{Ticker}_{i,t}$	0.08	−0.07	−0.53	0.01	0.22	0.07
	(0.22)	(−0.19)	(−1.56)	(0.05)	(1)	(0.31)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	214310	214208	214105	214002	213899	213799
Panel B: Filtered Ticker						
	(1)	(2)	(3)	(4)	(5)	(6)
	$Commit_{i,t}$	$Commit_{i,t+1}$	$Commit_{i,t+2}$	$Commit_{i,t+3}$	$Commit_{i,t+4}$	$Commit_{i,t+5}$
$Abn\ Attention^{Mention}_{i,t}$	−0.23	−0.94*	−0.3	−0.12	−0.12	0.19
	(−0.7)	(−1.72)	(−0.55)	(−0.51)	(−0.42)	(0.66)
$Abn\ Attention^{Official}_{i,t}$	6.53***	−7.44***	−6.69***	−1.11*	−0.16	1.07*
	(5.1)	(−5.85)	(−4.64)	(−1.72)	(−0.29)	(1.77)
$Abn\ Attention^{Ticker}_{i,t}$	−0.09	0.24	−0.56	0.02	0.19	0.23
	(−0.22)	(0.63)	(−1.44)	(0.08)	(0.72)	(0.75)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	140771	140702	140633	140563	140493	140426

6 Conclusion

Given the size of the cryptocurrency market and the importance of Twitter as a source of information, studying their link is essential for understanding the cross-section of cryptocurrency expected returns. Instead of considering only one dimension of Twitter activity, our paper is the first academic study as our knowledge to consider several samples of tweets. Distinguishing among the different types of tweets is important, as they have different functionalities, reach and sources. Consistent with Benedetti and Kostovetsky (2021), we show that Twitter impacts the cross-section of cryptocurrency expected returns through an overreaction channel (Barber & Odean, 2008; Da et al., 2011). By using several Twitter samples, we show however that this overreaction channel is mainly due to tweets posted by users and not by cryptocurrencies. Our results emphasize the heterogeneity of social media content, highlighting the need for researchers and practitioners to carefully consider which types of social media contents best suit their needs. Given the impact of popular tweets on future Twitter attention and it exacerbating effect on behavioral biases, we provide evidence that even small influential users can profit from their popularity to manipulate cryptocurrency prices.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 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*(1), 33–72.

Auer, R., Cornelli, G., Doerr, S., Frost, J., & Gambacorta, L. (2023). Crypto trading and bitcoin prices: evidence from a new database of retail adoption. *Unpublished Working Paper*.

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*(6), 2224–2276.

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

Banerjee, S. (2011). Learning from prices and the dispersion in beliefs. *Review of Financial Studies*, *24*(9), 3025–3068.

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*(2), 785–818.

Barillas, F., & Shanken, J. (2017). Which alpha? *Review of Financial Studies*, *30*(4), 1316–1338.

Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can twitter help predict firm-level earnings and stock returns? *Accounting Review*, *93*(3), 25–57.

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

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

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

Cao, S., Jiang, W., Yang, B., & Zhang, A. L. (2020). How to talk when a machine is listening?: Corporate disclosure in the age of ai. *Unpublished Working Paper*.

Carlin, B. I., Longstaff, F. A., & Matoba, K. (2014). Disagreement and asset prices. *Journal of Financial Economics*, *114*(2), 226–238.

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*(5), 1367–1403.

Chordia, T., Goyal, A., & Saretto, A. (2020). Anomalies and false rejections. *Review of Financial Studies*, *33*(5), 2134–2179.

Chui, A. C., Subrahmanyam, A., & Titman, S. (2022). Momentum, reversals, and investor clientele. *Review of Finance*, *26*(2), 217–255.

Colla, P., & Mele, A. (2010). Information Linkages and Correlated Trading. *Review of Financial Studies*, *23*(1), 203–246.

Cong, L. W., Karolyi, G. A., Tang, K., & Zhao, W. (2021). Value premium, network adoption, and factor pricing of crypto assets. *Unpublished Working Paper*.

Cong, L. W., Li, Y., & Wang, N. (2021). Tokenomics: Dynamic adoption and valuation. *Review of Financial Studies*, *34*(3), 1105–1155.

Cookson, J. A., Engelberg, J. E., & Mullins, W. (2023). Echo chambers. *Review of Financial Studies*, *36*(2), 450–500.

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? evidence from a social network of investors. *Journal of Finance*, *75*(1), 173–228.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, *66*(5), 1461–1499.

Du, J., Huang, D., Liu, Y.-J., Shi, Y., Subrahmanyam, A., & Zhang, H. (2022). Retail investors and momentum. *Unpublished Working Paper*.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, *75*(3), 1327–1370.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 1121–1152.

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

Hillert, A., Jacobs, H., & Müller, S. (2014). Media makes momentum. *Review of Financial Studies*, *27*(12), 3467–3501.

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

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.

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*.

Jiang, L., Liu, J., Peng, L., & Wang, B. (2022). Investor attention and asset pricing anomalies. *Review of Finance*, *26*(3), 563–593.

Kogan, S., Makarov, I., Niessner, M., & Schoar, A. (2023). Are cryptos different? evidence from retail trading. *Unpublished Working Paper*.

Lins, K. V., Roth, L., Servaes, H., & Tamayo, A. (2021). Sexism, culture, and firm value: Evidence from the harvey weinstein scandal and the# metoo movement. *Unpublished Working Paper*(2019-509).

Liu, Y., Sheng, J., & Wang, W. (2022). 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*(6), 2689–2727.

Liu, Y., Tsyvinski, A., & Wu, X. (2022). Common risk factors in cryptocurrency. *Journal of Finance*, *77*(2), 1133–1177.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance*, *66*(1), 35–65.

Lyandres, E., Palazzo, B., & Rabetti, D. (2022). Initial coin offering (ico) success and post-ico performance. *Management Science*, *68*(12), 8658–8679.

Mai, F., Bai, Q., Shan, J., Wang, X. S., & Chiang, R. H. (2015). The impacts of social media on bitcoin performance. *Unpublished Working Paper*.

Naeem, M. A., Mbarki, I., Suleman, M. T., Vo, X. V., & Shahzad, S. J. H. (2021). Does twitter happiness sentiment predict cryptocurrency? *International Review of Finance*, *21*(4), 1529–1538.

Pedersen, L. H. (2022). Game on: Social networks and markets. *Journal of Financial Economics*, *146*(3), 1097–1119.

Petersen, M. A. (2008). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, *22*(1), 435–480.

Pursiainen, V., & Toczynski, J. (2022). Retail investors' cryptocurrency investments. *Unpublished Working Paper*.

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

Shams, A. (2019). What drives the covariation of cryptocurrency returns. *Unpublished Working Paper*.