How are Social and Mass Media Different in Relation to the Stock Market? A Study on Topic Coverage and Predictive Value[1]

Hang Dong [a], Jie Ren [b,*], Balaji Padmanabhan [c], Jeffrey V. Nickerson [d]

[a] *IE University, Spain, Calle Cardenal Zu˜niga 12, 40003 Segovia, Spain* [b] *Gabelli School of Business, Fordham University, 140 West 62nd Street, New York, NY 10023, United States of America* [c] *Information Systems & Decision Sciences, Director, Center for Analytics & Creativity, Muma College of Business, University of South Florida, 4202 E Fowler Ave, Tampa, FL 33620, United States of America* [d] *School of Business, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, United States of America*

Abstract

Although investors in financial markets have access to information from both mass media and social media, trading platforms that curate and provide this information have little to go by in terms of understanding the difference between these two types of media. This paper compares social media with mass media in the stock market, focusing on information coverage diversity and predictive value with respect to future stock absolute returns. Based on a study of nearly a million stock-related news articles from the Sina Finance news platform and 12.7 million stock-related social media messages from the popular Weibo platform in China, we find that social media covers less stocks than mass media, and this effect is amplified as the volume of media information increases. We find that there is some short-term predictive value from these sources, but they are different. Although mass media information coverage is more predictive than social media information coverage in a one-day horizon, it is the other way around in a two-to five-day horizon. These empirical results suggest that social media and mass media serve stock market investors differently. We draw connections to theories related to how crowds and experts differ and offer practical implications for the design of media-related IS systems.

1 Last author's copy. Published version: Dong, H., Ren, J., Padmanabhan, B., & Nickerson, J. V. (2022). How are social and mass media different in relation to the stock market? A study on topic coverage and predictive value. Information & Management, 59(2), 103588. https://doi.org/10.1016/j.im.2021.103588

Keywords: social media, mass media, information coverage diversity, predictive value, future stock absolute return

1. Introduction

In the last few years, trading platforms have started to incorporate social media feeds for investors in the stock market. Figure 1 presents an example from a Bloomberg terminal that shows what an investor might see today from these platforms, where representative headlines are typically from mass media[2]; links to Twitter are also provided on one of the tabs.

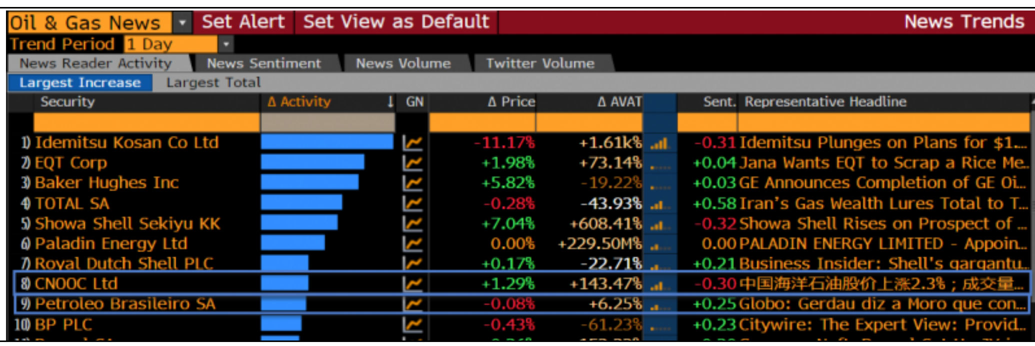

Figure 1. A Screenshot of Bloomberg Terminal Showing Media Information on Oil and Gas Stocks

As trading platforms, which currently favor mass media content, attempt to integrate the possibly noisy social media feeds, questions should be asked as to the reason and the value of doing so. In particular, how different are these two types of information and how can they help investors make trading decisions in the context of the stock market? If information from social media simply

[2] Here, we define mass media as the media outlets that are traditional and run by organizations, such as Wall Street Journal, where news articles are written by professional journalists and curated by editors.

replicates that from mass media or mainly adds noise, then information integration may even hurt investors. This question can only be answered if we understand the information differences between mass media and social media in the specific context of the stock market. Building on a unique and integrated dataset of both social media data and curated financial news from traditional media this paper aims to provide answers about the effects of media on investor decision-making, and, more generally, about the differences between these forms of media as information sources. These differences can not only inform theory but also have practical implications for news platform designers and for news platform users.

Specifically, understanding the information difference between social and mass media can help retail investors make more informed decisions. In particular, in the height of the recent pandemic many millennials rushed into the US stock market, relying on social media as their main source of information.[3] Free trading platforms, such as Robinhood, influence their decision-making and their stock trading.[4] The recent rise and fall of "meme" stocks highlights the importance of understanding the quality of the social media information in relationship to mass media news.[5] In combination, misapprehension or misweighting of media information can lead to not only personal risk but systemic risk. By contrast, understanding the differences between social and mass media can result in more intentionally designed information systems in general, and better designed financial information systems in particular [1,2]. Trading systems such as Bloomberg can carefully design to integrate diverse types of information to aid investor decision-making, as compared to ad-hoc "tab-based" presentation of information which essentially punts this responsibility to investors to decide if and how to use which type of information.

[3] https://www.statista.com/statistics/1010456/united-states-millennials-news-consumption/
[4] https://www.cnn.com/2020/06/12/investing/millennials-investing-robinhood/index.html
[5] https://www.forbes.com/advisor/investing/gamestop-meme-stocks-bb-amc-nok/

Although comparison of different media types can be done using different lenses, in this paper we compare the *diversity of information coverage* and *predictive value* of social media and mass media information. We focus on information diversity - the extent to which information is not homogenous [3]. Information diversity has been studied in prior literature with mixed implications, in part because of the human element involved. Some have indeed suggested that diverse information can inform investors and increase market efficiency [4]. Others suggest diverse information may cause information overload and information anxiety [5], which may hurt investors' decision-making. Because of such debate, our paper focuses on information diversity to study the informativeness of social and mass media information.

We argue there are two forces on social media that affect information diversity. The force driving convergence is herding [6,7], the force driving divergence is balkanization [8,9]. On the one hand, because there are many individual users of social media making decisions about what to discuss, without central control, there is, just as with a physical crowd, a tendency toward herding. That is, individual bloggers will take signals from each other, and blog about what is already being blogged about. This can lead to a focus on few rather than many topics. On the contrary, many social media bloggers are citizen journalists [8], so that social media functions like a discussion board that collects the diverse information bloggers generate [9]. Similar to the way the Internet facilitates research balkanization [9], social media, because it empowers individual perspectives, can create balkanization, generating the fragmentation of information. Both herding and balkanization in social media have been noted by previous scholars [10–12]. Our paper tests which force, herding or balkanization, is dominant on social media in the stock market, resulting in different levels of information diversity.

The nature of social media contrasts with the nature of mass media where information creators are professionals. Mass media professionals have a defined role: they cover news topics in all possible sections and thereby generate the information in all possible sections, producing fragmentation. Arguably, on mass media, herding is less likely to happen and balkanization is more likely to occur.

We study the two mechanisms, herding and balkanization, that affect information diversity on social versus mass media, and use the mechanisms as a way to compare the two types of media. In the stock market, information diversity measures how many stocks are being discussed and how many times they are discussed, providing a way to gauge the stream of information that affects the investors. Specifically, to measure information diversity, we define the information coverage of one stock as the number of social media posts (or mass media news articles) about the stock as a fraction of all social media posts (or mass media news articles) about any stock. We also examine the predictive value of information, which is the association between the information coverage and the future stock absolute return. In particular, we ask two questions:

1) How is the information coverage of social media different from that of mass media?
2) What is their relative value in terms of predicting short-term stock absolute returns?

We conducted our analyses with a unique and massive integrated dataset of financial news, microblogs and stock price data. We obtained stock-relevant social media microblogs from Sina Weibo, a social media platform similar to Twitter. Weibo is considered the most influential microblogging platform in China [13]. Weibo is also one of the largest social media platforms in the world, partly because of its dominant market position in China. We used Sina Finance, an Internet news platform in China (similar to Yahoo! Finance), to obtain mass media news stories published by mass media professionals or organizations in the same time period. Sina Finance is one of the most important mass media news aggregators that present all the news articles from all mass media outlets in China. Our integrated dataset includes 12.7 million stock-related weibos and more than 0.7 million pieces of stock news for 2014.

We found that social media serves as a filter to reduce information diversity: in the stock market the information coverage of social media is much *less* diverse than that of mass media, and that this information diversity difference is amplified when the volume of media information increases. In other words, we documented evidence for the herding mechanism that triggers social media users to select a few topics instead of all topics to cover, and mass media, on the contrary, is less subject to herding. Also, we found that the filtered subset of information has more predictive power in a relatively longer run than mass media information in general: even though mass media (which is less subject to herding) has a higher predictive value for stock absolute returns than social media in a one-day horizon, in a 2- to 5- day (or to 10 day) horizon social media (which is more subject to herding) is still predictive, whereas mass media is not. Generally speaking, our findings suggest that both types of information (herded (social media) versus non-herded (mass media)) do have different short-term signals in predicting absolute returns, both of which are of value to investors.

Our paper makes significant contributions to both research and practice. First, our findings show evidence that the dominant force that drives a lack of information diversity on social media is herding. Furthermore, increased information volume amplified this effect. One important difference between our paper and other studies on media is the following. Prior studies focused on social media users' behaviors and thus the herding behavior itself [14,15]. Our paper, in contrast, argues that information coverage diversity and predictive value are associated with the consequences of herding, and explores whether herding is beneficial or detrimental to information diffusion on social media in the stock market context. Second, our findings revisit the comparison between crowd-based models and expert-based models in the field of information generation and quality [16,17]. Our findings also suggest that social media and mass media complement each other in terms of information diversity and predictive value in with respect to stock returns. Thus, understanding both types of media information is crucial for investors, as well as the designers of the trading platforms.

On the one hand, investors can make more informed investment decisions given both social and mass media information. On the contrary, trading platforms should provide both types of information, but should explore more thoughtful design choices to help investors stay well-informed while making good trading decisions.

2. Background: Media in the Stock Market

Prior literature has extensively studied mass media and its role in the financial markets [18,19]. There are three possible mechanisms that prior literature has examined regarding how mass media impacts markets. The first mechanism treats mass media as a proxy for *investor sentiment* [19–21]. This stream of literature assumes that social media can represent collective opinions on stocks. Social media information is considered to measure collective mood states and can be used to associate with not only the stock market performance indicators such as Dow Jones Industrial Average (DJIA), NASDAQ and S&P 500 [22,23], but also individual stock performance [24,25]. The second mechanism treats mass media information coverage as a proxy for *attention* focused on stocks or the stock market. That is, investors tend to trade stocks that are discussed in the mass media [26–28]. The last mechanism treats mass media as a channel to gradually diffuse *fundamental information* among investors, which affects prices [29–31]. This data-driven approach has used mass media text to predict market reactions [32].

There has been some recent work that has considered *both* social and mass media together in looking at the impact on markets [33,34]. However, the paucity of literature in the intersection highlights the void here, and the opportunity for a systematic investigation into the relative content and value of mass and social media in the financial markets. That is, trading systems are being designed in a knowledge vacuum with respect to this intersection, and thereby make assumptions that are incomplete at best.

Table 1 provides a high-level summary of the main ideas in this section. It shows that there is substantial research that has looked at both media and the market. It also shows there is a lack of systematic insight into the conceptual differences between mass and social media in the markets—a role that this paper attempts to provide.

Table 1. Literature Summary	
Focus of the Study	*Example Papers*
Mass Media	
Finance Literature	
Proxy for Investor Sentiment/Nonfundamental	Tetlock (2007)[19]; Dougal et al. (2012)[20]; Tetlock (2011)[21]
Proxy for Investor Attention	Barber and Odean (2008)[26]; Yuan (2015)[27]; Solomon et al. (2014)[28]
Information Channel	Tetlock (2010)[30]; Tetlock et al. (2008)[31]; Peress (2014)[29]; Fang and Peress (2009)[35]
Data-driven Approach	Groß-Klußmann and Hautsch (2011) [32]
IS Literature	
Text-mining	Schumaker and Chen (2009)[36]; Hagenau et al. (2013)[37]; Feuerriegel and Prendinger (2016)[38]; Ming et al. (2014)[39]; Li et al. (2014)[34]; Hu et al. (2018)[40]
Social Media	
Finance Literature	
Investor Sentiment/Nonfundamental	Antweiler and Frank (2004)[41]; Sprenger et al. (2014)[42]; Siganos et al (2014)[43]
Fundamental Information Channel	Chen et al (2014)[44]
Data-driven Approach	Das and Chen (2007)[45]
IS Literature	
Predictive/Explanatory Power	Bollen et al. (2011)[23]; Luo et al (2013)[46]; Tirunillai and Tellis (2012)[47]; Deng et al. (2018)[24]; Li et al. (2018) [48]
Information Environment	Xu and Zhang (2013)[49]
Text Mining or Sentiment Analysis Techniques	Li et al. (2014)[25]; Nguyen et al. (2015) [50]
Mass Media and Social Media	
Finance Literature	
Contents in Social Media and Mass Media	Jiao et al (2020)[33]
IS Literature	

Combined Effect	Li et al. (2014) [25]

Specifically, our paper aims to study how social media and mass media are different from each other. A handful of studies in general contexts have examined the content differences between social media and mass media [16,51]. For example, Greenstein and Zhu[16] showed that contents in crowd-based models (e.g., Wikipedia) and expert-based models (e.g., Encyclopedia Britannica) present different biases and slants. Those authors argued that "little is known about how well collective decision-making performs along other dimensions of information quality, such as objectivity, consistency, relevance, and timeliness." Along these lines, to our best knowledge, our paper is among the first attempts to fill this gap by studying the dimensions of information coverage and its predictive value with respect to stock returns.

3. Herding versus Balkanization on Social Media and Mass Media

3.1 Information Diversity

Information quality is crucial to investors—investors desire information that is unbiased and can describe the whole picture. Thus, information diversity - the extent that information is not homogenous [3] - is one important aspect of the quality of information. This construct has been studied in different settings and has shown many effects [3,52,53]. On the one hand, some scholars point out its positive impacts. For example, in the finance literature, Goldstein and Yang[3] found the diversity of information can increase the amount of information that prices reveal. In the entrepreneurship literature, Hull et al. [53] showed that using information diversity a start-up is more likely to have a born-global strategy and at the same time is less likely to quit. In the innovation literature, there has been consensus on the positive impact of information diversity on innovation. The diversity of crowd members' backgrounds contributes to the benefits of crowdsourcing [54].

On the contrary, information diversity can increase information overload and distract people from making correct and timely decisions [55].

As information quality is important to stock investors, our paper examines the level of information diversity on social media versus mass media and explores the impact of their respective levels of information diversity on stock returns.

3.2 Herding or Balkanization on Social Media

Social media has been known to have a herding effect [6,7], which drives the homogeneity of information and reduces its diversity. For example, Li and Wu[6] found that the information of "daily deals" on Groupon facilitates herding, which increases sales. Trinkle et al. [7] found that exposure to other social media users' comments can facilitate herding toward similar reactions and perceptions to the news. Relevant to the stock market context, Sabherwal et al. [15] explicitly argued that an online stock message board is essentially a "herding device". Specifically, the two distinct characteristics of social media may facilitate the herding process. First, social media is quick in generating and diffusing information [56]. Social media users, particularly information creators, are able to observe what information is posted by others before creating information. This exposure is likely to lead to herding behavior, a phenomenon in which the observation of the majority leads observers to convert their views to match the consensus, regardless of their private information [57–59]. As social media consists of a large group of connected people that are exposed to each other's posts [8,60–63], the herding effect can be strong. Second, social media is essentially unregulated and so has more freedom to report whatever social media users want to read. So social media users do not have assigned specializations that limit their freedom to publish whatever they want in their tweets or weibos [56], facilitating herding further.

There is another characteristic of social media to be considered that may have an opposing effect. Information on social media comes from individual social media users. These users, functioning as citizen journalists [8], can contribute content anywhere anytime, which may increase the diversity of information. This is called balkanization—a process of fragmentation where information does not leave its own pocket. Balkanization has been mainly studied in the geopolitical settings. For example, Frey (1996)[64] studied demographic balkanization that happened due to immigration and domestic migration in the US. Specific to the IS domain, Alstyne and Brynjolfsson (1996) [9] showed the Internet facilitates research balkanization across geographic boundaries. In another paper [65] they explored whether information technologies can unite communities or balkanize communities. Welch (2010) [66] found that Wikipedia is also not immune to balkanization. In addition, scholars studied cyberbalkanization in the settings of online micro-lending and Facebook networking [67,68]. In our studied context, balkanization on social media can happen because social media users essentially are citizen journalists, reporting news events surrounding them. At the same time, many users also cite and discuss news articles on social media, creating another channel to increase information diversity on social media.

Therefore, on social media, two forces, herding and balkanization, can potentially affect information diversity, and may push it in different directions.

3.3 Herding versus Balkanization on Mass Media

Mass media, on the contrary, has media professionals creating news content. Mass media editors mainly select and report news based on their own professional judgment [56]. In addition, mass media professionals are less likely to be exposed to each other's news article before deciding what to write for publication, and their choice of news articles to write tends to be independent from each other. Therefore, mass media is less likely to be subject to herding behavior with respect to information generation. Also, mass media professionals have their defined specializations that in aggregate cover news topics in a wide range of sections, which in turn leads to fragmentation of information. In this way, mass media is more likely to be subject to balkanization with respect to information generation.

3.3 Herding versus Balkanization on Social Media and Mass Media in the Stock Market

In sum, on social media, it can be argued that herding and balkanization both may affect information diversity in different directions, with herding decreasing diversity and balkanization increasing diversity. Our paper aims to test which of the two forces is dominant in social media. Prior literature suggests that, at least in the context of financial markets, herding is likely to be the dominant force [15,69,70]. As Shleifer and Summers (1990) [71] note, individuals "tend to make the same mistake; they do not make random mistakes". Social media mainly represents the crowd's choice of information and thus is likely to be subject to herding with respect to stock market news. Balkanization, on the contrary, is not really studied with respect to investment behaviors; it is only studied by a few scholars at higher regulatory and macro-economic levels [72]. In sum, we argue that herding is the dominant force on social media and aim to test this assertion.

With respect to mass media, the professional journalists' defined specializations lead to covering news topics in a wide range of categories, thus having a higher likelihood of increasing the diversity of information and producing balkanization. Table 2 summarizes herding versus balkanization on social and mass media in the stock market.

Table 2 Herding versus Balkanization on Social and Mass media in the Stock Market

	Herding	Balkanization
Social Media	More	Less
Mass Media	Less	More

This leads to the following hypothesis.

H1. Information coverage of social media is less diverse than that of mass media in the stock market.

Specific to the stock market, the volume of information varies over time, and sometimes can be quite "spiky". This typically happens when there are unexpected shocks to the stock market such as

Brexit and the increase of interest rate. In the next hypothesis, we examine how the information coverage diversity in both social and mass media changes as the volume of both social and mass media information changes. This question is important, especially with respect to the design of the trading platforms. During the shocks, much information is generated and diffused in the market. As more information is collected by trading platforms, what information should trading platforms present and/or highlight for investors to use? And would the diversity of information coverage of either media still differ or become similar? The answers to these questions can contribute to the previous literature where media plays a key role in affecting investors' behaviors [73–75]. Investors may overreact to the shocks for some stocks and underreact for other stocks [76]. Our findings can help investors understand the information coverage diversity dynamics as the information volume changes (as a result of the shocks) and further make their investment decisions.

We argue that as media information volume increases, the herding effect tendency is different between social and mass media. The herding effect suggests that the larger the *N* (the number of social media posts for example), the more apparent the effect is [77]. One possible reason is that when the number of social media posts increases, it is more convenient for social media users to be exposed to others' posts. This can lead to an even more skewed distribution of information coverage, which means even less diverse information coverage on social media. In comparison, mass media publishes information independently and is also more subject to contextual constraints as discussed above. Therefore, if their information volume increases, their tendency to publish a small selection of topics compared to that of social media is still low. Thus, we conjecture:

H2. When information volume increases, the information coverage of social media is even less diverse than that of mass media in the stock market.

4. Predictive Values of Social Media Information Coverage versus Mass Media Information Coverage in Predicting Absolute Stock Returns

H1 and H2 together conjecture that compared to mass media, social media functions as a filter to filter out a subset of information. Now the natural question to ask is: is herding of social media information beneficial or detrimental for people's decision-making? Typically, it is difficult to measure information quality in settings such as the political context or the retail market as each individual has his or her own definitions of what is better. Fortunately, in the stock market, there is a clear and consensual criterion—whether the media's information has a better predictive value in predicting future stock returns [78]. That is, we are able to ask and test the following question. Which media's information coverage—social media (with herded information) versus mass media (with non-herded information) — can have a higher predictive power in the stock market? Findings to this question can also significantly contribute to the prior literature (in Table 1) on the predictive or explanatory value of media information in the stock market.

Some readers may argue that this question is less important as investors can read information from both mass media and social media. We, nevertheless, hold a different view for two reasons. First, investors, particularly individual investors, are subject to limited attention [26]. If they can only manage to read and comprehend a subset of information, then what information to select is crucial, and knowing the value of each can help investors (and trading platforms) make better decisions on what information to consider. Second, even if AI technologies are able to read all information and present all information to investors, the question regarding how investors should use the potential noisy social media information in conjunction with mass media information is still open.

We assume that obtaining stock returns is the most important objective for stock retail investors, who are the typical information users in this context. The impact of media coverage on returns can be positive and negative. For example, intensive media coverage with positive (negative) sentiment predicts positive (negative) future returns [24]. To avoid this empirical problem, we use the absolute value of return as our dependent variable. By doing so, we change the complex relationship between the absolute value of return and media coverage into a monotonically increasing or decreasing shape. Here, we assume that information related to positive returns and negative returns are equally important for investors. That is, investors may sell stocks to avoid negative returns or short sell stocks to gain positive returns in the presence of future negative returns. Therefore, we focus on the absolute value of future stock returns to test and compare both social and mass media coverage's predictive power in the stock market context. To be clear, this study is not about the causal relationships between the variables, but is instead about the predictive value of the variables. We note that previous studies have documented the causal relationship between social media and stock returns [24,79] and between mass media and stock returns [20,29].

Given the lack of detailed theory supporting either direction, we propose to examine both possibilities:

H3. Social media has a higher (lower) predictive value than mass media in predicting future stock absolute returns.

5. Data and Method

We integrated two large datasets to answer the questions. The first dataset is from Sina Weibo. Sina Weibo is an equivalent of Twitter and is considered as the most influential social media platform in China. Sina extracts weibos that mention Chinese stocks by using ticker and *jiancheng*, short name in Chinese. For example, "600000 Pufa Bank" represents a listed company "Pufa Bank". The combination of the ticker and the *jiancheng* typically does not have any meaning in daily life. By doing so, we could perfectly match the stocks and the weibos. The dataset includes nearly 12.7 million weibos for 3,410 stocks in 2014.

We calculated **social media information coverage** by the ratio of the number of weibos that mention stock *i* on day *t* to the number of weibos on any stock on day *t*. We denoted it as $PerSM_{i,t}$:

$$PerSM_{i,t} = \frac{NumSM_{i,t}}{NumSM_t}$$

where $NumSM_{i,t}$ is the number of weibos that mentioned stock *i* on day *t* and $NumSM_t$ is the number of weibos on any stock on day *t*. By doing so, we changed our dataset from cross-sectional data at the weibo level to panel data at the stock-day level.

We used a dataset from Sina Finance to measure mass media financial news. Sina Finance is one of the most important news aggregators of public stock news in China (similar to Yahoo Finance). Our news data includes almost all publicly available news about the Chinese stock market in 2014. We dropped news articles that mentioned more than one stock because those news articles were more likely to be associated with the stock market than individual stocks. This dataset includes 748,013 articles of stock news that mentioned one stock. We defined **mass media information coverage**, $PerMM_{i,t}$, as the ratio of the number of news articles that mentioned stock *i* on day *t* to the number of news articles on any stock on day *t*:

$$PerMM_{i,t} = \frac{NumMM_{i,t}}{NumMM_t}$$

where $NumMM_{i,t}$ is the number of news articles that mentioned stock *i* on day *t* and $NumMM_t$ is the number of news articles on any stock on day *t*.

We used the Gini coefficient to measure **information diversity**. The Gini coefficient was originally designed to measure inequality [80,81] and is higher when information is less diverse (more focused). Following Foster et al.[80] and Lokshin and Sajaia[82], we defined the Gini coefficient as follows:

$$Gini\_weibo_t = \frac{\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} |PerSM_{i,t} - PerSM_{j,t}|}{2N_t \sum_{i=1}^{N_t} PerSM_{i,t}}$$

$$Gini\_news_t = \frac{\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} |PerMM_{i,t} - PerMM_{j,t}|}{2N_t \sum_{i=1}^{N_t} PerMM_{i,t}}$$

where N_t is the number of stocks in the stock market on day *t*; and $Gini\_weibo_t$ and $Gini\_news_t$ are Gini coefficients on day *t* for social media and mass media, respectively. In other words, for every day in our dataset we calculated the Gini coefficient of information coverage for social media and the Gini coefficient for mass media, respectively.

We defined information volume as the following.

$$Vol\_weibo_t = \ln(1 + Num\_weibo_t)$$

$$Vol\_news_t = \ln(1 + Num\_news_t)$$

Num_weibo_t is the number of weibos in the stock market on day *t*; Num_news_t is the number of news articles in the stock market on day *t*.

The proxy for the predictive value of media is the association between information coverage and the absolute value of future stock return, which is consistent with the concept of "informativeness" in the stock market that has been used in the finance literature [78]. We conducted regressions with absolute return as the dependent variable and the information coverage as the independent variable to calculate its coefficient as a proxy for the predictive value of media. We used the absolute value of raw return for stock *i* on day *t*, denoted by $|r_{i,t}|$, as the predicting objective, which is **future stock absolute returns** [78]. Table 3 lists the definitions of all variables. Table 4 provides the summary statistics, and Table 5 shows the correlations.

Table 3. Definitions of Variables	
PerSM{*i,t*}	Number of weibos on stock *i* on day *t* over number of weibos on day *t*
PerMM{*i,t*}	Number of news articles on stock *i* on day *t* over number of news articles on day *t*
$Gini\_weibo_t$	The Gini coefficient of the information coverage distribution on social media on day *t* [80,82]
$Gini\_news_t$	The Gini coefficient of the information coverage distribution on mass media on day *t* [80,82]
Vol_weibo_t	Natural logarithm of one plus the number of weibos on any stock on day *t*
Vol_news_t	Natural logarithm of one plus the number of news on any stock on day *t*
$\|r_{i,t}\|$	The absolute value of raw return for stock *i* on day *t*.

Table 4. Summary Statistics
Panel 1: Day Level (H1 and H2)
N = 364

	Mean	Std. Dev.	Median	Min	Max
Gini_weibo{t}	0.6989	0.1446	0.7276	0.2338	0.9041
Gini_news{t}	0.1866	0.0645	0.1605	0.0652	0.3965
Vol_weibo{t}	9.1688	2.3688	9.9788	1.7918	12.1979
Vol_news{t}	7.0194	1.4857	7.9735	2.4849	8.0709
Panel 2: Stock-Day Level (H3)					
N = 429,257					
Per_SM{i,t}	0.0003	0.0018	0.0001	0.0000	0.1873
Per_MM{i,t}	0.0004	0.0004	0.0003	0.0000	0.0325
|r_{i,t}|	0.0165	0.0153	0.0122	0.0000	0.0997
Vol_weibo{t}	10.5661	0.7592	10.5214	7.7376	12.1979
Vol_news{t}	7.9941	0.0271	7.9960	7.8789	8.0709

Table 5. Correlations

Note: *p< 0.10, **p<0.05, ***p<0.01

Panel 1: Day Level (H1 and H2)					
	Gini_weibo{t}	Gini_news{t}	Vol_weibo{t}	Vol_news{t}	
Gini_weibo{t}	1.0000				
Gini_news{t}	-0.4478***	1.0000			
Vol_weibo{t}	0.6939***	-0.7010***	1.0000		
Vol_news{t}	0.6373***	-0.7665***	0.8841***	1.0000	
Panel 2: Stock-Day Level (H3)					
	Per_SM{i,t}	Per_MM{i,t}	|r_{i,t}|	Vol_weibo{t}	Vol_news{t}
Per_SM{i,t}	1.0000				
Per_MM{i,t}	0.3691***	1.0000			
|r_{i,t}|	0.0907***	0.0084***	1.0000		
Vol_weibo{t}	0.0000	0.0000	0.0707***	1.0000	
Vol_news{t}	0.0000	0.0000	-0.0348	0.8841***	1.0000

6. Results

To test our first hypothesis, we compared the Gini coefficient of social media information coverage with that of mass media information coverage via a paired samples *t*-test. Social media's Gini efficient was significantly larger than that of mass media (0.6989 versus 0.1866; t = 53.4530; p

< 0.001), suggesting that social media's information coverage on average is less diverse than that of mass media. Therefore, H1 was supported.

Table 6. Comparison Between Information Coverage Diversity of Social Media versus Mass Media					
	Obs	Mean	Std. Dev.	Diff	t
Gini_weibo	364	0.6989	0.1446	0.5124	53.4530
Gini_news	364	0.1866	0.0645		

We then generated the kernel density plot to show the distributions of Gini coefficients of social media versus mass media over the year of 2014 in our dataset. We also plotted the Gini coefficient means of both distributions (Figure 2). Consistent with Table 6, we found that the Gini coefficient of social media information coverage is typically larger than that of mass media.

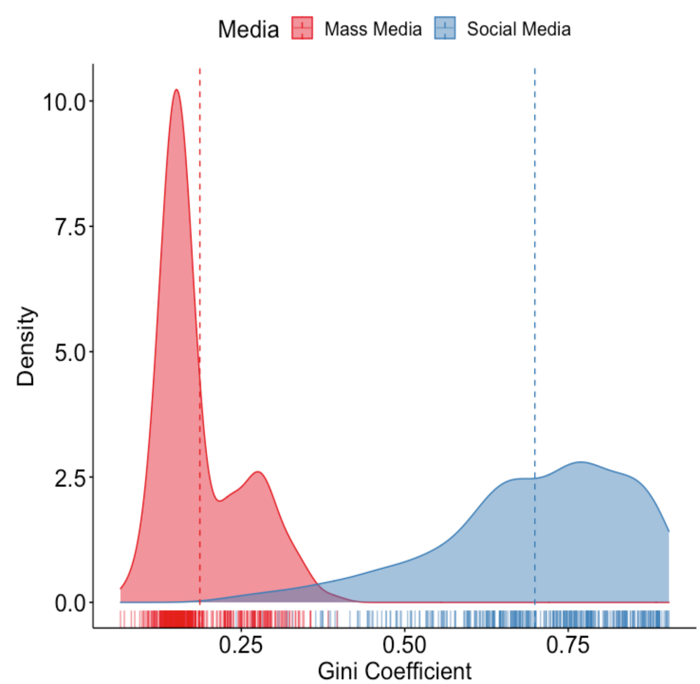

Figure 2. Comparison Between Gini Coefficients of Information Coverage of Social Media versus Mass Media

To test H2, we regressed the Gini coefficient of social media information coverage (that of mass media information coverage), *Gini_weibo* (*Gini_news*), at *t* on the number of weibos (that of news articles), *Vol_weibo* (*Vol_news*), at *t*, respectively. We also controlled for the lagged coefficients for social media as well as for mass media, which are *Gini_weibo* and *Gini_news*, from *t-1* to *t-5*.

Columns 1 and 2 in Table 7 show that the volume of information is positively associated with the Gini coefficient of information coverage on social media (0.0407, $p < 0.01$; 0.0296, $p < 0.01$), suggesting an increase in the volume of social media posts is associated with *less* diversity of information coverage on social media. In contrast, columns 3 and 4 in Table 7 show that the volume of information is negatively associated with the Gini coefficient of information coverage on mass media (-0.0230, $p < 0.01$; -0.0235, $p < 0.01$), suggesting an increase in the volume of mass media news article is associated with *more* diversity of information coverage on mass media. We interpret the results as when the information volume increases, the information coverage difference between social media and mass media increases. Therefore, H2 was supported.

Table 7. Information Coverage Diversity Results When Information Volume Increases
Note: The standard errors are reported in parentheses. *p<0.10; **p<0.05; ***p<0.01

DV	Gini_weibo{t}		Gini_news{t}	
	(1)	(2)	(3)	(4)
Vol_Weibo{t}	0.0407*** (0.0033)	0.0296*** (0.0033)		
Vol_News{t}			-0.0230*** (0.0023)	-0.0235*** (0.0023)
Gini_weibo{t-1}		0.3698*** (0.0501)		
Gini_weibo{t-2}		-0.0460 (0.0534)		
Gini_weibo{t-3}		0.0546 (0.0521)		
Gini_weibo{t-4}		0.0202 (0.0521)		
Gini_weibo{t-5}		0.0136 (0.0466)		
Gini_news{t-1}				-0.0258 (0.0463)
Gini_news{t-2}				0.1485*** (0.0456)
Gini_news{t-3}				0.0770 (0.0453)
Gini_news{t-4}				0.0963** (0.0452)
Gini_news{t-5}				0.0367

				(0.0445)
Day of Week Fixed Effects	Yes	Yes	Yes	Yes
Obs	364	359	364	359
Adjusted R squared	0.4844	0.5669	0.6394	0.6542

To test our last set of competing hypotheses, we regressed the absolute value of stock return on day *t* on social media information coverage from day *t-1* to day *t-5*, *PerSM{i,t-1}* to *PerSM{i,t-5}*, as well as traditional media information coverage from day *t-1* to day *t-5*, *PerMM{i,t-1}* to *PerMM{i,t-5}*, controlling for the absolute value of stock raw return on day *t-1*. Table 8 and Figure 3 report the results.

Table 8. Comparison Between the Predictive Values of Social Media versus Mass Media

Note: The standard errors are clustered at the stock level and day level. They are reported in parentheses. *p<0.10; **p<0.05; ***p<0.01

DV	|r_{i, t}|		
	(1)	(2)	(3)
PerSM{i,t-1}	0.2573*** (0.0457)	0.2532*** (0.0453)	0.2309*** (0.0461)
PerSM{i,t-2}			0.0470*** (0.0140)
PerSM{i,t-3}			0.0353*** (0.0095)
PerSM{i,t-4}			0.0246*** (0.0082)
PerSM{i,t-5}			0.0332*** (0.0093)
PerMM{i,t-1}		0.7423*** (0.1629)	0.6552*** (0.1570)
PerMM{i,t-2}			0.0162 (0.0349)
PerMM{i,t-3}			0.0156 (0.0195)
PerMM{i,t-4}			-0.0061 (0.0142)
PerMM{i,t-5}			-0.0043 (0.0177)
|r_{i, t-1}|	0.0374*** (0.0053)	0.0371*** (0.0053)	0.0365*** (0.0054)
Stock Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Obs.	429,257	429,257	426,960
Adjusted R squared	0.1579	0.1580	0.1589

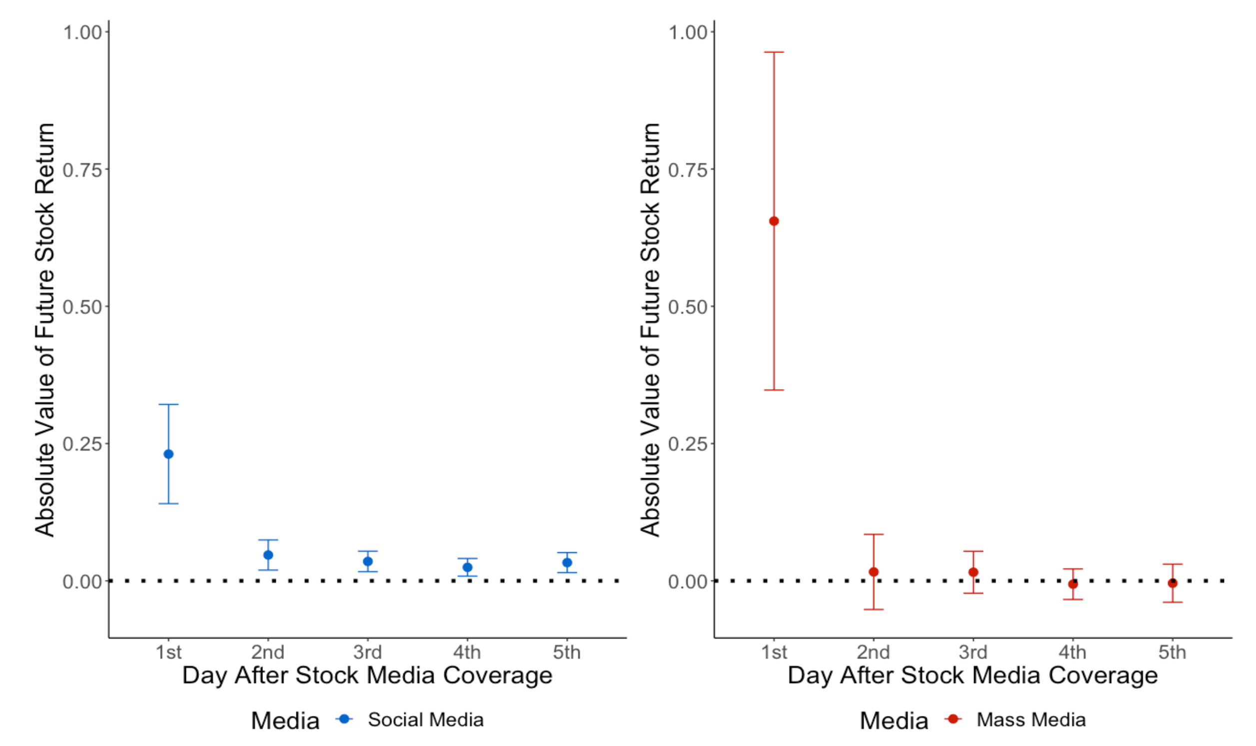

Figure 3. Predictive Value of Stock Media Coverage for Different Time Horizons

We focused on the coefficients of either social media information coverage or mass media information coverage to see how predictive either media is. Across the three models (columns 1 to 3 in Table 8), the coefficients of social media information coverage on day *t-1*, *PerSM{i,t-1}*, are positive and significant (0.2573, $p<0.01$; 0.2532, $p<0.01$; 0.2309, $p<0.01$). One percentage increase in the social media coverage may lead to around 0.25% increase in the absolute value of stock raw return. This suggests that social media is predictive in a one-day horizon. The predictive value of social media may simply result from mass media's impact—social media posts may simply repeat the information that occurs on mass media. Therefore, we controlled for mass media information coverage on day *t-1* in columns 2 and 3. The coefficient of *PerSM{i,t-1}* in either column 2 or column 3 is still positive and significant, suggesting that social media is predictive above and beyond mass media information coverage. Across columns 2 and 3, the coefficients of mass media information coverage on day *t-1* are positive and significant as well (0.7423, $p<0.01$; 0.6552, $p<0.01$). A one percentage shift in mass media coverage increases the absolute value of stock raw return by 0.74% and 0.66% (in columns 2 and 3). This suggests that mass media is also predictive in a one-day horizon.

We conducted two Wald tests to compare the magnitude of the coefficients of *PerSM*{*i,t-1*} and *PerMM*{*i,t-1*}. Rows 1 and 2 in Table 9 show the results. The coefficients of *PerSM*{*i,t-1*} are significantly smaller than those of *PerMM*{*i,t-1*}. Mass media information coverage is more predictive than social media information coverage in a one-day horizon.

Table 9. Wald Tests

	Null Hypothesis	F-statistics	P-value
Column 2 in Table 6	$\beta_{PerSM\{t-1\}} = \beta_{PerMM\{t-1\}}$	7.94	0.0053
Column 3 in Table 6	$\beta_{PerSM\{t-1\}} = \beta_{PerMM\{t-1\}}$	6.29	0.0130
Column 3 in Table 6	$\beta_{PerSM\{t-2\}} + \beta_{PerSM\{t-3\}} + \beta_{PerSM\{t-4\}} + \beta_{PerSM\{t-5\}} = \beta_{PerMM\{t-2\}} + \beta_{PerMM\{t-3\}} + \beta_{PerMM\{t-4\}} + \beta_{PerMM\{t-5\}}$	3.69	0.0563

In column 3 of Table 8, we controlled for 2- to 5-day lagged social media information coverage and mass media information coverage. The coefficients of *PerSM*{*i,t-1*} and *PerMM*{*i,t-1*} are similar to those in column 2. Interestingly, we found that the coefficients of social media information coverage on days from *t-2* to *t-5, PerSM*{*i,t-2*} to *PerSM*{*i,t-5*}, are positive and significant, whereas the coefficients of mass media information coverage on days from *t-2* to *t-5, PerMM*{*i,t-2*} to *PerMM*{*i,t-5*}, are not significant. We also conducted a Wald test to compare the coefficients of social media information coverage from *t-2* to *t-5* with those of mass media information coverage during the same time period. Row 3 in Table 9 shows the results. The sum of the coefficients of the 2- to 5-day lagged social media information coverage is significantly larger than that of mass media coverage during the same time period at the 10% level. This result suggests that social media is more predictive than mass media in a 2- to 5-day horizon.

We did not state H3 directionally since ex-ante it was not clear which media might be more predictive. However, after these results, it is worth asking: why? Why is it that both social and mass media have predictive power in the very short-term one-day horizon, whereas it is only social media that continues to have a more prolonged effect on return?

Although we do not offer a final word on this, we conjecture that this is may be due to several reasons. First, in the 2- to 5-day period, social media activity might be an attention proxy. Stocks with more attention (social media activity) might see larger trading behavior, leading to higher absolute value of the returns. Once there is a shock (social media or mass media news) that moves the stock, investors may subsequently use social media activity as a proxy for when interest in the stock wanes. Second, mass media is more associated with explicit/less noisy and fundamental information. Therefore, information on mass media can quickly be incorporated into the stock price in one day and has no association with future stock absolute returns in the subsequent days. In contrast, social media includes both noisy and informative messages. Information on social media is implicit, but has a persistent impact. But these are all interesting questions to explore in future work.

7. Further Analyses

7.1 Granger Causality Tests

It is not new that there is an information flow between social media and mass media. On the one hand, many media professionals post news articles using their own social media accounts. Sismeiro and Mahmood (2018)[83] showed that social media users use the news links that are diffused on social media to migrate to news websites for the detailed content of the news articles. At the same time, social media users also cite and discuss news articles, increasing the information flow from mass media to social media. Jiao et al (2020)[33] explicitly argue that "social media contents are generated by repeating and discussing (e.g., re-tweeting) existing news". On the contrary, mass media professionals pick up information from social media to report. Ren et al (2021)[84] show that mass media outlets may use social media to find what their readers want to read, and consequently change reporting tones.

In the stock market, the relationship among social media, mass media, and stock returns is complex. Information on social media and mass media can overlap—social media diffuses information from mass media and mass media responds to stock returns [33,84]. Meanwhile, social media and mass media also respond to past returns [24]. Therefore, the analyses in our previous section may not be sufficient to show their relationship. In this section, we conducted panel vector autoregressions (PVAR) and Granger Causality following Deng et al. (2018) and Dewan and Ramaprasad (2014)[24,85] to further analyze their complex relationship.

This PVAR model requires our variables to be non-intermittent. However, stock markets only open during the weekdays—during the weekends/holidays we have missing observations. Therefore, following Deng et al. (2018)[24], we combined the social (mass) media volume during the weekends with those on Mondays, and recalculated our variables at the stock-trading day level.

We first tested the stationarity of all our three variables using the Fisher-type Augmented Dickey-Fuller test following Choi (2001)[86]. All statistics are large enough to reject the null hypothesis that the variables contain unit roots. We then ran the PVAR model specified as follows:

$$\begin{bmatrix} |r_{i,t}| \\ PerSM_{i,t} \\ PerMM_{i,t} \end{bmatrix} = \sum_{j=1}^{J} \begin{bmatrix} {\beta_{11}}^{t-j} & {\beta_{12}}^{t-j} & {\beta_{13}}^{t-j} \\ {\beta_{21}}^{t-j} & {\beta_{22}}^{t-j} & {\beta_{23}}^{t-j} \\ {\beta_{31}}^{t-j} & {\beta_{32}}^{t-j} & {\beta_{33}}^{t-j} \end{bmatrix} \begin{bmatrix} |r_{i,t-j}| \\ PerSM_{i,t-j} \\ PerMM_{i,t-j} \end{bmatrix} + \begin{bmatrix} |\mu_{r,\mathrm{i},t}| \\ \mu_{SM,\mathrm{i},t} \\ \mu_{MM,\mathrm{i},t} \end{bmatrix}$$

Where J is the order of the model.

The optimal lag for the model is one day based on Bayesian information criterion (BIC). However, the BIC chooses the most efficient model to predict the dependent variable—some inefficient independent variables (e.g., *PerMM*{*i,t-2*} to *PerMM*{*i,t-5*}) may be discarded in the model. The purpose of the paper is to compare the predictive power of *PerSM* with that of *PerMM*. So, we need to estimate the predictive power of variables that are even not efficient. Accordingly, we report the PVAR results using five day lags as in our previous regressions, rather than one day lag suggested by the BIC. We used the Helmer transformation on the endogenous variables following Abrigo and Love (2016)[87]. The estimation method is GMM. The results of the estimations are reported in Table 10. Columns 1 and 2 in Table 10 report the estimation results when PerSM is the response. The coefficients of PerMM{i,t-1} are positive and statistically significant and the coefficients of PerMM from *t–3* to *t–5* are not significant and the coefficient of PerMM{i,t-2}is significant but negative. The results suggest that social media coverage follows mass media coverage at *t-1* to discuss news items, but the relationship is short-lived and only lasts for one day. Columns 3 and 4 report the results on how mass media coverage responds to social media coverage. The coefficients of PerSM at *t–1* are positive and statistically significant. The coefficients of PerSM at other day lags are not significant. The results show that mass media coverage responds to social media coverage in one-day-horizon.

We then analyzed how stock returns respond to social media coverage and mass media coverage. The literature also provides abundance of evidence showing that stock returns respond to both social media and mass media. For example, Tetlock (2007)[19] show that mass media sentiment is associated with subsequent stock returns. Dougal et al (2012)[20] further show the causality relationship between mass media and stock returns. On the contrary, social media also affects stock returns [24,79].

Columns 5 and 6 report our results. The coefficients of PerSM from *t-1* to *t-5* are statistically significant at 1% and positive. In contrast, the coefficients of PerMM at *t-2*, *t-4,* and *t-5* are insignificant and the coefficient of PerMM at *t-3* is significant but at 10%. The results are consistent with those in Section 6.

Finally, we conducted the Granger causality tests. Table 11 reports the results. The Granger causality tests show that mass (social) media coverage Granger-causes social (mass) media coverage.

And both mass media coverage and social media coverage Granger-causes the absolute value of returns.

Table 10. PVAR Estimation Results

Note: The standard errors are clustered at the stock level. They are reported in parentheses.
*p<0.10; **p<0.05; ***p<0.01

	(1)	(2)	(3)	(4)	(5)	(6)
	PerSM		PerMM		|r|	
PerSM{i,t-1}	0.1236*** (0.0274)	0.1379*** (0.0344)	0.0005 (0.0004)	0.0011*** (0.0004)	0.2409 *** (0.0359)	0.2263*** (0.0365)
PerSM{i,t-2}		0.0686*** (0.0158)		0.0002 (0.0003)		0.1313*** (0.0344)
PerSM{i,t-3}		0.0493*** (0.0107)		0.0000 (0.0004)		0.1435*** (0.0299)
PerSM{i,t-4}		0.0470*** (0.0143)		0.0000 (0.0004)		0.1065*** (0.0286)
PerSM{i,t-5}		0.0347*** (0.0142)		-0.0004 (0.0003)		0.0818*** (0.0250)
PerMM{i,t-1}	0.2037*** (0.0361)	0.1876*** (0.0338)	0.2770 *** (0.0229)	0.2576*** (0.0192)	1.0044 *** (0.1718)	0.8012*** (0.1474)
PerMM{i,t-2}		-0.0767*** (0.0171)		0.0517*** (0.0093)		0.1679 (0.1036)
PerMM{i,t-3}		-0.0222 (0.0217)		0.0345*** (0.0089)		0.1876* (0.0975)
PerMM{i,t-4}		-0.0109 (0.0199)		0.0410*** (0.0113)		0.0673 (0.0845)
PerMM{i,t-5}		0.0106 (0.0166)		0.0474*** (0.0077)		0.1988 (0.1292)
|r_{i, t-1}|	0.0058*** (0.0008)	0.0048*** (0.0009)	-0.0001 (0.0000)	-0.0001** (0.0001)	-0.0225 *** (0.0037)	-0.0544*** (0.0055)
|r_{i, t-2}|		0.0001 (0.0009)		-0.0002 (0.0001)		-0.0529*** (0.0057)
|r_{i, t-3}|		0.0002 (0.0008)		-0.0002*** (0.0001)		-0.0419*** (0.0055)
|r_{i, t-4}|		0.0003 (0.0008)		-0.0002** (0.0001)		-0.0485*** (0.0053)
|r_{i, t-5}|		0.0001 (0.0008)		-0.0001*** (0.0000)		-0.0428*** (0.0051)

Table 11. Granger Causality Test Results Note: The table reports the Chi-square of the Granger causality tests. *p<0.10; **p<0.05; ***p<0.01			
	PerSM	PerMM	|r|
PerSM		32.754**	45.207***
PerMM	59.811***		42.017***
|r|	231.752***	17.076***	

7.2 Further Evidence on Attention-driven Herding

As social media information coverage is less diverse than that of mass media, we further explored what stock information social media focuses on as compared to mass media—especially can these stocks be attention-grabbing stocks that may have attracted people to herd on social media? In other words, would these stocks be large-sized stocks or stocks that have large absolute stock returns? Tables 12 and 13 show our descriptive analysis. We categorized all stocks in the stock market from the lenses of size and absolute value of stock return. We used market capitalization as a proxy for stock size, which is calculated as the natural logarithm of one plus the market capitalization of the company stock i on day t. And consistent with our main analysis, we used the absolute value of stock raw return as another lens. For each lens, we ranked all the stocks into five portions with 5 as the largest value and 1 as the smallest value. And then based on stock and day, we matched the information coverage of social media with that of mass media and conducted paired t-tests. We found that social media tends to focus on larger-sized stocks and stocks with larger absolute value of stock returns to report than mass media (in Table 12: size (3): 0.0004006 > 0.0003674, p<0.001; size (4): 0.0004276 > 0.0003737, p<0.001; size (5): 0.0006589 > 0.0005471, p<0.001; in Table 13: return (5): 0.0005431 > 0.0003878, p<0.001). We also created one plot (Figure 4) for Tables 12 and 13. One possible explanation is that social media favors attention-grabbing stocks to cover as large stock size and large absolute value of stock returns are proxies for attention-grabbing stocks.

Table 12. Comparison of Information Coverage Between Social Media and Mass Media on Stock Size Note: *** p<0.001.						
Size	Social Media	Mass Media	Difference	Obs	t-statistics	P value
1 (smallest)	0.0002331	0.0003511	-0.0001180	85,582	-23.9903	***
2	0.0003168	0.0003594	-0.0000426	131,408	-8.7048	***
3	0.0004006	0.0003674	0.0000331	133,218	5.5805	***
4	0.0004276	0.0003737	0.0000539	119,745	9.1871	***
5 (Largest)	0.0006589	0.0005471	0.0001118	99,995	11.7839	***

Table 13. Comparison of Information Coverage Between Social Media and Mass Media on the Absolute Value of Stock Return Note: *** p<0.001.						
Absolute Value of Stock Return	Social Media	Mass Media	Difference	Obs	t-statistics	P value
1 (smallest)	0.0003028	0.0004078	-0.0001050	111,693	-16.9063	***
2	0.0003260	0.0003951	-0.0000691	113,583	-8.2117	***
3	0.0003204	0.0003815	-0.0000611	110,344	-9.2309	***
4	0.0003432	0.0003753	-0.0000321	111,396	-4.2404	***
5 (Largest)	0.0005431	0.0003878	0.0001553	111,133	17.1219	***

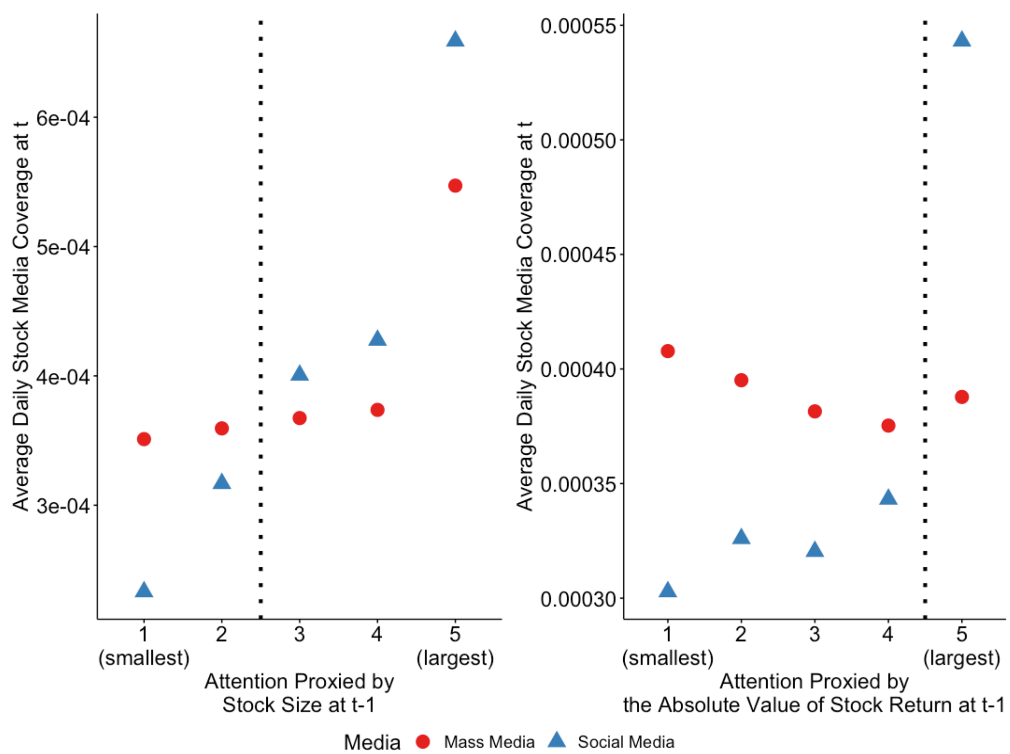

Figure 4. Stocks' Characteristics and Stock Media Coverage

8. Robustness Checks

We conducted four robustness checks. All our findings and main results remain the same. First, we included more time lags in the regressions for our third hypothesis to examine for a longer time (from *t-2* to *t-10*) if social media information coverage is persistently predictive, but mass media information coverage is not. Here we found that social media has a higher predictive value than mass media in a two- to ten-day horizon (see Table A.1 in the Appendix).

Second, we used the absolute value of raw return in testing our third hypothesis. However, a common approach in the finance literature is also to use the abnormal return. In a robustness check, we followed the finance literature and used the absolute value of abnormal return, estimated by the Fama–French–Carhart four factor model [88]. Table A.2 reports the results.

Third, one concern is that some social media publishers are media professionals. For example, MarketWatch, a professional news and analysis provider, also has a social media account (@MarketWatch). To address this concern, we excluded all professional users in the social media. In particular, some Weibo users can voluntarily reveal their professional identities (e.g., financial analysts, mutual fund managers, or media professionals) to the public. Sina Weibo labels them as V-users. In this check, we excluded social media messages published by verified users (they account for 19.65% of our overall social media messages). Consequently, we compared purely social media messages, without the influence of mass media channels on social media, with mass media news. Tables A.3 through A.5 show the results.

Fourth, another concern is that our results on the predictive power (H3) can be driven by the stock market performance. In particular, the CSI300 index skyrocketed by 51.63% in 2014. To relieve this concern, we reran the regression using the sample from January 2014 to June 2014. During this period, the stock market performance dropped by 7% - the bull market only happened after June 2014. Table A.6 reports the results, which are consistent with our previous results.

9. Discussion

Our study explores the information differences between social and mass media. Understanding the fundamental differences between social and mass media are critical for investors in the stock market, as well as designers of platforms.

9.1 Theoretical Implications

From a theoretical perspective, our findings make three important contributions to the prior literature. First and foremost, in the stock market we discover social media functions as a filter to narrow information for investors to consider. Furthermore, increased information volume amplified this effect. Our findings may relate to previous work on homophily and herding. Some studies [14,89] have shown that social media can facilitate homophily, and that the social influence facilitated by social media [90] leads to group think. Thus herding emerges. Susarla et al. [91] used homophily as one of the mechanisms to explain the diffusion of user generated content. In the finance literature, Sabherwal et al. [15] showed that social media is a herding device. Our study suggests mass media is less subject to this mechanism, perhaps due to media professionals' domain knowledge and professional and rational judgement, or perhaps due to mass media's regulation.

Other mechanisms may also interact with this mechanism to reinforce social media's herding capacity. In our robustness check, we found that the stocks that social media focuses on to cover tend to be attention-grabbing. In other words, the tendency to be attracted only to attention-grabbing topics may constrain social media from covering a wider range of topics, whereas mass media may be able to overcome such a constraint—possibly due to regulation and the media professionals' work ethic that encourages them to impartially report topics worth reporting.

Our findings suggest that social media filters appear to work in favor of short-term investors. The psychological limitations (herding and the tendency to be attracted to attention-grabbing topics) that social media is subject to did not necessarily hurt the quality of the information that social media has focused on to cover. Social media had predictive value and was even more predictive in the (relatively) longer run than mass media in the context of the stock market. Especially in the subsequent days (from *t-2* to *t-5* in Table 7), social media was more predictive than mass media in being associated with future stock returns. These findings are not inconsistent with the literature that casts doubt on the information quality of social media [92–94], but provides nuance by presenting evidence in support of social media having predictive value as time goes on.

Second, our paper contributes to the new stream of literature that compares social media with mass media in the stock market. We found that even in a utilitarian context, such as the stock market, where people are supposed to be rational, social media's information coverage focuses on a small number of stocks and has predictive values that are different than mass media. That is, our paper revisits the comparison between crowd-based models and expert-based models in the field of information generation and quality in the stock market [16]. We focus on the dimensions of information diversity and predictive value to compare social media with mass media in the stock market. Our findings add to the literature that suggests the crowd-based models are superior and can function as the prediction market [95–97].

Third, specific to the stock market context, our findings add to studies that use media information to predict or explain future stock returns. Previous literature did not treat social media and mass media as different information sources; instead, it focused on one media at a time and uniformly theorized information from either social or mass media into different categories, including fundamentals, non-fundamentals, and investors' mood states to justify how information affects future stock returns [19,30,41]. We argue that scholars may also need to consider the structural difference in the coverage information from social versus mass media to potentially further theorize and differentiate social/mass media.

9.2 Practical Implications

First, our findings could help trading platforms to provide tools that dissuade individuals from making poor investing decisions. Our findings show that both social media and mass media have predictive values and these values are different and may complement each other. In the stock market context that we studied, social media filters out a few topics to focus on, whereas mass media has a higher diversity of topics to cover, thus social media and mass media complement each other in terms of topic coverage. Also, social media is better at predicting the absolute value of future stock returns in the relatively longer run (from $t+2$ to $t+10$), and mass media has a stronger predictive value for the next day, thus these two types of media may complement each other in terms of their impact on people's decisions. Therefore, news recommender systems should incorporate both types of media and accordingly give readers a more holistic picture of information.

Second, although investors can be offered customization to their preferences, the importance of good defaults are now generally understood [98]. Our results suggest the need for such default trading screens and interfaces to combine social media and mass media. Specific to the stock market setting, traders may have different goals of trading: short term or long term. The designers of the relevant information systems may need to match the information sources (social versus mass) with traders' different goals when presenting them with relevant information. For example, traders with the short-term goal may be presented with more mass media information and traders with the relatively long-term goal may be presented with more social media information.

10. Conclusions and Future Work

This paper systematically compared social media with mass media on the dimensions of information coverage diversity and predictive value in the financial market. Our findings have practical and theoretical implications and offer numerous opportunities for further contributions.

There are many directions for future work. Our findings for instance implicitly suggest that herding is the dominant force in social media. Future research might seek to detect each process and differentiate these forces, with the goal of discovering causal mechanism that might be at work. In particular, future research might test the conjecture that mass media and social media may work together as filtering and amplifying mechanisms that eventually lead to herding as information is disseminated back and forth.

Future research might also seek to generalize our findings, first beyond the Chinese stock market to other markets, and then more broadly to industries other than finance. For example, the entertainment industry often takes the advantage of the word of mouth of the fan base and accordingly exploits their idols' commercial value. What the fan base wants to read is considered as valuable information—fans with Bieber fever always desire more information about Justin Bieber. In these markets, the perceived value of information by readers (here social media users) almost equals the real value of information. We argue our findings regarding the information filtering mechanism of social media can be more apparent and lead to more informative consequences in these contexts, but need more detailed study in future research.

There is also a need for caution suggested by our findings. Social media's information filtering mechanism is highly subject to the manipulation that builds up the collective attention for certain information topics. Although we might expect mass media to dampen manipulation, there may be cases where it fails to do so. In recent examples of the recent "meme" stock phenomena, the distribution of information is highly skewed. One consequence is that any speculator could create a trending topic on a new meme stock by publishing either true or false information to obtain substantial profits. Future work might need to look into that.

To test H3, we assumed that positive returns and negative returns are equally important for retail investors. Even though both positive and negative returns are both important, the equal relationship that we assumed may not always hold true. Barber and Odean (2008) show evidence that retail investors are relatively less likely to short sell stocks [26]. That is, in the context we studied, all investors can buy stocks but only investors who are current shareholders can sell them—in this context, all investors are sensitive to positive media coverage and only shareholders are sensitive to negative media coverage. Future research might include sentiment in an analysis to explore how positive coverage and negative coverage of stocks on social media versus mass media have different predicting values.

We do not claim causality in H3. An intriguing interpretation of our results is causal—information flows in social media and mass media lead to the price jumps. In Section 7.1, we showed that media coverage granger-causes changes in stock returns. We call for future studies that can provide stronger evidence for the causality relationship between media coverage and stock returns.

Although our dataset is at the daily level, these results might also hold on high frequency data. This choice is natural because most mass media, particularly newspapers, make major updates once per day. However, it is both important and interesting to explore how social media coverage varies across time using higher frequency data. Such future research might examine this in anticipation that even newspapers are shifting to more frequent updates on their web sites.

Acknowledgement

The authors are indebted to Yuanping WANG, Xikun SUN, Lu LIU and Sina Corporation, who provided the data. This material is based upon work supported by the National Science Foundation under grants 2128906, 2113906, and 1909803.

References

[1] P. Gomber, R.J. Kauffman, C. Parker, B.W. Weber, On the fintech revolution: interpreting the forces of innovation, disruption, and transformation in financial services, J. Manag. Inf. Syst. 35 (2018) 220–265.

[2] P. Gomber, R.J. Kauffman, C. Parker, B.W. Weber, Financial Information Systems and the Fintech Revolution, (2018).

[3] I. Goldstein, L. Yang, Information diversity and complementarities in trading and information acquisition, J. Finance. 70 (2015) 1723–1765.

[4] S. Grossman, On the efficiency of competitive stock markets where trades have diverse information, J. Finance. 31 (1976) 573–585.

[5] D. Bawden, L. Robinson, The dark side of information: overload, anxiety and other paradoxes and pathologies, J. Inf. Sci. 35 (2009) 180–191.

[6] X. Li, L. Wu, Herding and Social Media Word-of-Mouth: Evidence from Groupon, Ssrn. (2013). https://doi.org/10.2139/ssrn.2264411.

[7] B.S. Trinkle, R.E. Crossler, F. Bélanger, Voluntary disclosures via social media and the role of comments, J. Inf. Syst. 29 (2015) 101–121.

[8] S.C. Lewis, K. Kaufhold, D.L. Lasorsa, Thinking about citizen journalism: The philosophical and practical challenges of user-generated content for community newspapers, Journal. Pract. 4 (2010) 163–179.

[9] M. Van Alstyne, E. Brynjolfsson, Could the Internet balkanize science?, Science (80-.). 274 (1996) 1479–1480.

[10] K. Sasahara, W. Chen, H. Peng, G.L. Ciampaglia, A. Flammini, F. Menczer, Social influence and unfollowing accelerate the emergence of echo chambers, J. Comput. Soc. Sci. (2020) 1–22.

[11] P.M. Leonardi, M. Huysman, C. Steinfield, Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations, J. Comput. Commun. 19 (2013) 1–19.

[12] L. Rose, T. Bartoli, Agnotology and the epistemology of ignorance: a framework for the propagation of ignorance as a consequence of technology in a Balkanized media ecosystem, Postdigital Sci. Educ. 2 (2020) 184–201.

[13] E. Harwit, The Rise and Influence of Weibo (Microblogs) in China, Asian Surv. 54 (2014) 1059–1087. https://doi.org/10.1525/as.2014.54.6.1059.

[14] X. Li, L. Wu, Herding and social media word-of-mouth: Evidence from Groupon, Forthcom. MISQ. (2018).

[15] S. Sabherwal, S.K. Sarkar, Y. Zhang, Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news, J. Bus. Financ. Account. 38 (2011) 1209–1237. https://doi.org/10.1111/j.1468-5957.2011.02258.x.

[16] S. Greenstein, F. Zhu, Do experts or crowd-based models produce more bias? Evidence from

encyclopedia britannica and wikipedia, MIS Q. 42 (2018) 945–959. https://doi.org/10.25300/MISQ/2018/14084.

[17] J. Gunaratne, L. Zalmanson, O. Nov, The Persuasive Power of Algorithmic and Crowdsourced Advice, J. Manag. Inf. Syst. 35 (2018) 1092–1120.

[18] P.C. Tetlock, Information transmission in finance, Annu. Rev. Financ. Econ. 6 (2014) 365–384. https://doi.org/10.1146/annurev-financial-110613-034449.

[19] P.C. Tetlock, Giving content to investor sentiment : The role of media in the stock market, J. Finance. 62 (2007) 1139–1168.

[20] C. Dougal, J. Engelberg, D. Garcia, C.A. Parsons, Journalists and the stock market, Rev. Financ. Stud. 25 (2012) 639–679.

[21] P.C. Tetlock, All the news that's fit to reprint: Do investors react to stale information?, Rev. Financ. Stud. 24 (2011) 1481–1512. https://doi.org/10.1093/rfs/hhq141.

[22] X. Zhang, H. Fuehres, P.A. Gloor, Predicting stock market indicators through twitter "I hope it is not as bad as I fear," Procedia-Social Behav. Sci. 26 (2011) 55–62.

[23] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, J. Comput. Sci. 2 (2011) 1–8.

[24] S. Deng, Z. (James) Huang, A.P. Sinha, H. Zhao, The interaction between microblog sentiment and stock returns: An empirical examination, MIS Q. 42 (2018) 895–918. https://doi.org/10.25300/MISQ/2018/14268.

[25] X. Li, H. Xie, L. Chen, J. Wang, X. Deng, News impact on stock price return via sentiment analysis, Knowledge-Based Syst. 69 (2014) 14–23.

[26] B.M. Barber, T. Odean, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, Rev. Financ. Stud. 21 (2008) 785–818.

[27] Y. Yuan, Market-wide attention, trading, and stock returns, J. Financ. Econ. 116 (2015) 548–564.

[28] D.H. Solomon, E. Soltes, D. Sosyura, Winners in the spotlight: Media coverage of fund holdings as a driver of flows, J. Financ. Econ. 113 (2014) 53–72. https://doi.org/10.1016/j.jfineco.2014.02.009.

[29] J. Peress, The media and the diffusion of information in financial markets: Evidence from newspaper strikes, J. Finance. 69 (2014) 2007–2043.

[30] P.C. Tetlock, Does public financial news resolve asymmetric information?, Rev. Financ. Stud. 23 (2010) 3520–3557. https://doi.org/10.1093/rfs/hhq052.

[31] P.C. Tetlock, M. Saar-Tsechansky, S. Macskassy, More than words: Quantifying language to measure firms' fundamentals, J. Finance. 63 (2008) 1437–1467.

[32] A. Groß-Klußmann, N. Hautsch, When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions, J. Empir. Financ. 18 (2011) 321–340.

[33] P. Jiao, A. Veiga, A. Walther, Social media, news media, and the stock market, J. Econ. Behav. Organ. 176 (2020) 63–90.

[34] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, Y. Chen, The effect of news and public mood on stock movements, Inf. Sci. (Ny). 278 (2014) 826–840.

[35] L. Fang, J. Peress, Media Coverage and the Cross-Section of Stock Returns, J. Finance. 64 (2009) 2023–2052.

[36] R.P. Schumaker, H. Chen, Textual analysis of stock market prediction using breaking financial news: The AZFin text system, ACM Trans. Inf. Syst. 27 (2009) 12.

[37] M. Hagenau, M. Liebmann, D. Neumann, Automated news reading: Stock price prediction based on financial news using context-capturing features, Decis. Support Syst. 55 (2013) 685–697.

[38] S. Feuerriegel, H. Prendinger, News-based trading strategies, Decis. Support Syst. 90 (2016) 65–74.

[39] F. Ming, F. Wong, Z. Liu, M. Chiang, Stock market prediction from WSJ: text mining via sparse matrix factorization, in: 2014 IEEE Int. Conf. Data Min., IEEE, 2014: pp. 430–439.

[40] Z. Hu, W. Liu, J. Bian, X. Liu, T.-Y. Liu, Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction, in: Proc. Elev. ACM Int. Conf. Web Search Data Min., ACM, 2018: pp. 261–269.

[41] W. Antweiler, Z.M. Frank, Is all that talk just noise? the information content of Internet stock message boards, J. Finance. 59 (2004) 1259–1294. https://doi.org/10.1017/CBO9781107415324.004.

[42] T.O. Sprenger, A. Tumasjan, P.G. Sandner, I.M. Welpe, Tweets and trades: The information content of stock microblogs, Eur. Financ. Manag. 20 (2014) 926–957.

[43] A. Siganos, E. Vagenas-Nanos, P. Verwijmeren, Facebook's daily sentiment and international stock markets, J. Econ. Behav. Organ. 107 (2014) 730–743.

[44] H. Chen, P. De, Y. Hu, B.H. Hwang, Wisdom of crowds: The value of stock opinions transmitted through social media, Rev. Financ. Stud. 27 (2014) 1367–1403. https://doi.org/10.1093/rfs/hhu001.

[45] S.R. Das, M.Y. Chen, Yahoo! for Amazon: Sentiment extraction from small talk on the web, Manage. Sci. 53 (2007) 1375–1388. https://doi.org/10.1287/mnsc.1070.0704.

[46] X. Luo, J. Zhang, W. Duan, Social media and firm equity value, Inf. Syst. Res. 24 (2013) 146–163.

[47] S. Tirunillai, G.J. Tellis, Does chatter really matter? Dynamics of user-generated content and stock performance, Mark. Sci. 31 (2012) 198–215.

[48] T. Li, J. van Dalen, P.J. van Rees, More than just noise? Examining the information content of stock microblogs on financial markets, J. Inf. Technol. 33 (2018) 50–69.

[49] S.X. Xu, X.M. Zhang, Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investor Reaction, MIS Q. 37 (2013) 1043-A10. https://doi.org/10.1080/09668136.2012.671567.3.

[50] T.H. Nguyen, K. Shirai, J. Velcin, Sentiment analysis on social media for stock movement prediction, Expert Syst. Appl. 42 (2015) 9603–9611.

[51] N.I. Mondragon, L. Gil de Montes, J. Valencia, Ebola in the public sphere: a comparison between

mass media and social networks, Sci. Commun. 39 (2017) 101–124.

[52] S. Figlewski, Information diversity and market behavior, J. Finance. 37 (1982) 87–102.

[53] C.E. Hull, Z. Tang, J. Tang, J. Yang, Information diversity and innovation for born-globals, Asia Pacific J. Manag. (2019) 1–22.

[54] D.C. Brabham, Crowdsourcing, Mit Press, 2013.

[55] M.I. Hwang, J.W. Lin, Information dimension, information overload and decision quality, J. Inf. Sci. 25 (1999) 213–218.

[56] N. Newman, W.H. Dutton, G. Blank, Social media in the changing ecology of news: The fourth and fifth estates in Britain., Int. J. Internet Sci. 7 (2012) 6–22.

[57] A. V. Banerjee, A simple model of herd behavior, Q. J. Econ. 107 (1992) 797–817. https://doi.org/10.2307/2118364.

[58] R.J. Shiller, Conversation, information, and herd behavior, Am. Econ. Rev. 85 (1995) 181–185. https://doi.org/10.2307/2117915.

[59] S. Bikhchandani, D. Hirshleifer, I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades, J. Polit. Econ. 100 (1992) 992–1026. https://doi.org/10.1086/261849.

[60] N. Thurman, Forums for citizen journalists? Adoption of user generated content initiatives by online news media, New Media Soc. 10 (2008) 139–157.

[61] S.B. Bullard, Editors use social media mostly to post story links, Newsp. Res. J. 36 (2015) 170–183.

[62] J. Van Dijck, The Culture of Connectivity: A Critical History of Social Media, Oxford University Press, 2013.

[63] N. Newman, The rise of social media and its impact on mainstream journalism, 2009.

[64] W.H. Frey, Immigration, domestic migration, and demographic balkanization in America: new evidence for the 1990s, Popul. Dev. Rev. (1996) 741–763.

[65] M. Van Alstyne, E. Brynjolfsson, Global village or cyber-balkans? Modeling and measuring the integration of electronic communities, Manage. Sci. 51 (2005) 851–868.

[66] C. Welch, Measuring Balkanization in Wikipedia, (2010).

[67] C. Chan, K. Fu, Facebook pages' strength of ties and segregation of engaged users: An examination of the hierarchical model of cyberbalkanization, in: Proc. 50th Hawaii Int. Conf. Syst. Sci., 2017.

[68] G. Burtch, S. Wattal, A. Ghose, An empirical examination of cultural biases in interpersonal economic exchange, in: Int. Conf. Inf. Syst. ICIS 2012, 2012: pp. 3329–3346.

[69] F. Caparrelli, A.M. D'Arcangelis, A. Cassuto, Herding in the Italian stock market: a case of behavioral finance, J. Behav. Financ. 5 (2004) 222–230.

[70] I. Medhioub, M. Chaffai, Islamic finance and herding behavior: an application to Gulf Islamic stock

markets, Rev. Behav. Financ. (2018).

[71] A. Shleifer, L. Summers, The Noise Trader Approach to Finance, J. Econ. Perspect. 4 (1990) 19–33. https://doi.org/10.1257/jep.4.2.19.

[72] A. Coley, US Regulation of Cross-Border Banks: Is It Time to Embrace Balkanization in Global Finance, Va. J. Int'l L. 56 (2016) 701.

[73] D. Hirshleifer, S.S. Lim, S.H. Teoh, Driven to distraction: Extraneous events and underreaction to earnings news, J. Finance. 64 (2009) 2289–2325.

[74] A. Ben-Rephael, Z. Da, R.D. Israelsen, It depends on where you search: Institutional investor attention and underreaction to news, Rev. Financ. Stud. 30 (2017) 3009–3047. https://doi.org/10.1093/rfs/hhx031.

[75] J. Engelberg, R.D. Mclean, J. Pontiff, Anomalies and news, J. Finance. 73 (2018) 1971–2001. https://doi.org/10.1111/jofi.12718.

[76] J. Li, J. Yu, Investor attention, psychological anchors, and stock return predictability, J. Financ. Econ. 104 (2012) 401–419. https://doi.org/10.1016/j.jfineco.2011.04.003.

[77] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science (80-.). 286 (1999) 509–512.

[78] D. Bradley, J. Clarke, S. Lee, C. Ornthanalai, Are analysts' recommendations informative? Intraday evidence on the impact of time stamp delays, J. Finance. 69 (2014) 645–673. https://doi.org/10.1111/jofi.12107.

[79] H. Dong, J. Gil-Bazo, Sentiment stocks, Int. Rev. Financ. Anal. 72 (2020) 101573. https://doi.org/https://doi.org/10.1016/j.irfa.2020.101573.

[80] J. Foster, S. Seth, M. Lokshin, Z. Sajaia, A Unified Approach to Measuring Poverty and Inequality, Washington, DC, 2013.

[81] C. Gini, Variabilità e mutabilità, Rome: Libreria Eredi Virgilio Veschi (1955), 1912.

[82] M. Lokshin, Z. Sajaia, EGEN_INEQUAL: Stata module providing extensions to generate inequality and poverty measures, 2006.

[83] C. Sismeiro, A. Mahmood, Competitive vs. complementary effects in online social networks and news consumption: A natural experiment, Manage. Sci. 64 (2018) 4967–5460.

[84] J. Ren, H. Dong, B. Padmanabhan, J. V. Nickerson, Emergent Sentiment in Financial Markets: Sentiment Flow from Social Media to Mass Media, J. Assoc. Inf. Sci. Technol. (2021).

[85] S. Dewan, J. Ramaprasad, Social media, traditional media, and music sales., Mis Q. 38 (2014) 101–121.

[86] I. Choi, Unit root tests for panel data, J. Int. Money Financ. 20 (2001) 249–272. https://doi.org/10.1016/S0261-5606(00)00048-6.

[87] M.R.M. Abrigo, I. Love, Estimation of panel vector autoregression in Stata, Stata J. 16 (2016) 778–804. https://doi.org/10.1177/1536867x1601600314.

[88] M.M. Carhart, On persistence in mutual fund performance, J. Finance. 52 (1997) 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

[89] B. Gu, P. Konana, R. Raghunathan, H.M. Chen, Research note—The allure of homophily in social media: Evidence from investor responses on virtual communities, Inf. Syst. Res. 25 (2014) 604–617.

[90] L. Muchnik, S. Aral, S.J. Taylor, Social Influence Bias: A Randomized Experiment, Science (80-.). 341 (2013) 647–651.

[91] A. Susarla, J.-H. Oh, Y. Tan, Social networks and the diffusion of user-generated content: Evidence from YouTube, Inf. Syst. Res. 23 (2012) 23–41.

[92] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2017) 211–236.

[93] A. Zubiaga, M. Liakata, R. Procter, G.W.S. Hoi, P. Tolmie, Analysing how people orient to and spread rumours in social media by looking at conversational threads, PLoS One. 11 (2016) e0150989.

[94] A. Benham, E. Edwards, B. Fractenberg, L. Gordon-Murnane, C. Hetherington, D.A. Liptak, M. Smith, C. Thompson, A.P. Mintz, Web of deceit: Misinformation and manipulation in the age of social media, Information Today, Inc., 2012.

[95] H. Nishikawa, M. Schreier, S. Ogawa, User-generated versus designer-generated products: A performance assessment at Muji, Int. J. Res. Mark. 30 (2013) 160–167. https://doi.org/10.1016/J.IJRESMAR.2012.09.002.

[96] S.E. Page, Diversity and Complexity, Princeton University Press, Princeton, NJ, 2011.

[97] M.K. Poetz, M. Schreier, The value of crowdsourcing: Can users really compete with professionals in generating new product ideas?, J. Prod. Innov. Manag. 29 (2012) 245–256. https://doi.org/10.1111/j.1540-5885.2011.00893.x.

[98] S. Benartzi, R. Thaler, Heuristics and biases in retirement savings behavior, J. Econ. Perspect. 21 (2007) 81–104.

Appendix

Table A.1. H3 Results (With time lags from *t-1* to *t-10*) Note: The standard errors are clustered at the stock level and the day level. They are reported in parentheses. *p<0.10; **p<0.05; ***p<0.01	
DV	|r_{i, t}|
PerSM{i,t-1}	0.2817*** (0.0556)
PerSM{i,t-2}	0.0568*** (0.0177)
PerSM{i,t-3}	0.0310*** (0.0097)
PerSM{i,t-4}	0.0259*** (0.0098)
PerSM{i,t-5}	0.0130 (0.0083)
PerSM{i,t-6}	0.0463*** (0.0166)
PerSM{i,t-7}	0.0446** (0.0182)
PerSM{i,t-8}	0.0553*** (0.0164)
PerSM{i,t-9}	0.0129* (0.0071)
PerSM{i,t-10}	0.0109* (0.0056)
PerMM{i,t-1}	1.2174*** (0.2207)
PerMM{i,t-2}	0.0079 (0.0269)
PerMM{i,t-3}	0.0098 (0.0228)
PerMM{i,t-4}	0.0093 (0.0163)
PerMM{i,t-5}	0.0180 (0.0120)
PerMM{i,t-6}	0.0048 (0.0170)
PerMM{i,t-7}	-0.0286 (0.0243)
PerMM{i,t-8}	-0.0734* (0.0318)
PerMM{i,t-9}	0.0168 (0.0166)
PerMM{i,t-10}	-0.0021 (0.0104)
|r_{i, t-1}|	0.0777*** (0.0056)
Stock Fixed Effects	Yes
Time Fixed Effects	Yes
Obs	404,217
Adjusted R squared	0.1382

Table A.2. H3 Results (With the absolute value of abnormal return as DV as an alternative) Note: The standard errors are clustered at the stock level and the day level. They are reported in parentheses. *p<0.10; **p<0.05; ***p<0.01			
DV	|α_{i, t}|		
	(1)	(2)	(3)
PerSM{i,t-1}	0.3291*** (0.0568)	0.3215*** (0.0559)	0.2937*** (0.0554)
PerSM{i,t-2}			0.0633*** (0.0185)
PerSM{i,t-3}			0.0391*** (0.0101)
PerSM{i,t-4}			0.0315*** (0.0098)

PerSM{i,t-5}			0.0258*** (0.0085)
PerMM{i,t-1}		1.3244*** (0.2399)	1.2394*** (0.2233)
PerMM{i,t-2}			0.0005 (0.0270)
PerMM{i,t-3}			0.0039 (0.0222)
PerMM{i,t-4}			0.0006 (0.0157)
PerMM{i,t-5}			0.0123 (0.0124)
|r_{i, t-1}|	0.0437*** (0.0053)	0.0772*** (0.0056)	0.0763*** (0.0056)
Stock Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Obs	415,540	415,540	413,256
Adjusted R squared	0.1350	0.1354	0.1361

Table A.3. H1 Results
(With social media users as non-verified users)

	Obs	Mean	Std. Dev.	Diff	t-stat
Gini_Weibo	364	0.7299	0.1480	0.5433	54.1730
Gini_News	364	0.1866	0.0645		

Table A.4. H2 Results
(With social media users as non-verified users)
Note: The standard errors are reported in parentheses.
*p<0.10; **p<0.05; ***p<0.01

DV	Gini_weibo{t}		Gini_news{t}	
	(1)	(2)	(3)	(4)
Num_Weibo{t}	0.0571*** (0.0025)	0.0496*** (0.0031)		
Num_News{t}			-0.0230*** (0.0023)	-0.0235*** (0.0023)
Gini_weibo{t-1}		0.2110*** (0.0451)		
Gini_weibo{t-2}		-0.0764* (0.0455)		
Gini_weibo{t-3}		0.0489 (0.0443)		
Gini_weibo{t-4}		-0.0174 (0.0442)		
Gini_weibo{t-5}		-0.0155 (0.0389)		
Gini_news{t-1}				-0.0258 (0.0463)
Gini_news{t-2}				0.1485*** (0.0456)
Gini_news{t-3}				0.0770*

				(0.0453)
Gini_news{t-4}				0.0963** (0.0452)
Gini_news{t-5}				0.0367 (0.0445)
Day of Week Fixed Effects	Yes	Yes	Yes	Yes
Obs	364	359	364	359
Adjusted R squared	0.7378	0.5669	0.6394	0.6542

Table A.5. H3 Results
(With social media users as non-verified users)
Note: The standard errors are clustered at the stock level and the day level. They are reported in parentheses.
*p<0.10; **p<0.05; ***p<0.01

DV	|r_{i, t}|		
	(1)	(2)	(3)
PerSM{i,t-1}	0.1968*** (0.0431)	0.1935*** (0.0427)	0.1730*** (0.0429)
PerSM{i,t-2}			0.0406*** (0.0125)
PerSM{i,t-3}			0.0311*** (0.0088)
PerSM{i,t-4}			0.0252*** (0.0076)
PerSM{i,t-5}			0.0335*** (0.0082)
PerMM{i,t-1}		0.7587*** (0.1639)	0.6738*** (0.1581)
PerMM{i,t-2}			0.0213 (0.0353)
PerMM{i,t-3}			0.0199 (0.0193)
PerMM{i,t-4}			-0.0082 (0.0139)
PerMM{i,t-5}			-0.0043 (0.0175)
|r_{i, t-1}|	0.0381*** (0.0053)	0.0377*** (0.0053)	0.0372*** (0.0053)
Stock Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Obs	428,887	428,887	426,590
Adjusted R squared	0.1565	0.1573	0.1576

Table A.6. H3 Results **(With observations from January 2014 to June 2014)** Note: The standard errors are clustered at the stock level and the day level. They are reported in parentheses. *p<0.10; **p<0.05; ***p<0.01	
DV	|r_{i, t}|
PerSM{i,t-1}	0.2379*** (0.0274)
PerSM{i,t-2}	0.0409** (0.0172)
PerSM{i,t-3}	0.0400*** (0.0088)
PerSM{i,t-4}	0.0297*** (0.0081)
PerSM{i,t-5}	0.0321*** (0.0094)
PerMM{i,t-1}	0.6888 (0.2045)
PerMM{i,t-2}	-0.0057 (0.0277)
PerMM{i,t-3}	-0.0131 (0.0198)
PerMM{i,t-4}	-0.0184 (0.0197)
PerMM{i,t-10}	-0.0116 (0.0222)
|r_{i, t-1}|	0.0150**
	(0.0072)
Stock Fixed Effects	Yes
Time Fixed Effects	Yes
Obs	205,138
Adjusted R squared	0.1789