Crowdsourced Investors' Recommendations and Stock Return Synchronicity

Mustabsar Awais[+]
Management School, University of Sheffield,
Conduit Road, Sheffield S10 1FL, United Kingdom

Junhong Yang[*]
SOAS School of Finance and Management,
Thornhaugh Street, Russell Square, London WC1H 0XG

Last Revised: February 2021

Abstract

We investigate whether disagreement on StockTwits provides firm-specific information. Using supervised machine learning approaches and a novel dataset, we predict investors' recommendations and measure disagreement among investors on StockTwits. Our findings suggest that an increase in investors' disagreement results in a drop in return synchronicity. The negative impact of investors' disagreement on return synchronicity suggests higher inflows of firm-specific information. In line with this view, we find that disagreement improves price informativeness by increasing the price leads of earnings. Further empirical evidence suggests that the negative impact of disagreement on return synchronicity is more pronounced for firms with less transparent information environments and higher salience on StockTwits.

Keywords: Disagreement; Price informativeness; Firm-specific Information; Heterogeneity; Salience

JEL Classification: G12; G14; G40; G41

[+] mawais2@sheffield.ac.uk (M. Awais)
[*] junhong.yang@soas.ac.uk (J. Yang)

It is the mark of an educated person to search for the same kind of clarity in each topic to the extent that the nature of the matter accepts it. For it is similar to expect a mathematician to speak persuasively or for an orator to furnish clear proofs! Each person judges well what they know and is thus a good critic of those things, [to be a critic] one must be educated about everything.

-- Aristotle, Nicomachean Ethics

1. Introduction

Disagreement among investors has emerged as the centerpiece of behavioral finance research in understanding its role to drive investors' trading in the financial markets (Karpoff 1986; Varian 1989; Kandel & Pearson 1995; Hong & Stein 1999). These studies provide corroborative evidence about the deviation from rational models and highlight the need to explore investor trading behaviors to unleash plausible evidence of variations from the rational models. However, there is mixed evidence about the role of investors' disagreement in financial markets. For example, Carlin *et al.* (2014) argue that greater disagreement is directly proportional to expected returns as investors could face higher uncertainty and adverse selection when a disagreement arises. This argument is consistent with Varian (1985); Harris and Raviv (1993), and Banerjee and Kremer (2010), who suggest that heterogeneous priors play an essential role in increasing disagreement, and consequently, require additional compensation and higher risk premiums. In contrast, Miller (1977) argues that the divergence of opinions in the market leads to lower risk premiums in the presence of short-sale constraints, and asset prices should reflect the valuation of optimists. Studies by Chen et al. (2002) and Yu (2011) found compelling evidence in favor of Miller's hypothesis. Either way, these studies highlight the critical role of disagreement in financial markets and how it affects asset prices.

Prior studies have mainly focused on investigating the predictive power of disagreement and its underlying behavioral explanations. However, little is known about the mechanisms by which disagreement among investors influences financial markets and whether disagreement among investors offers firm-specific information and contributes to the efficiency of capital allocation. Therefore, building on Hong and Stein (2007) theoretical framework and motivated by the role of disagreement in the financial markets, we explore whether disagreement among investors results in less return synchronicity, consequently providing more firm-specific information.

A critical challenge for researchers is to find a suitable proxy[1] for disagreement among investors. With the recent advances in technology, there has been a paradigm shift in how market participants consume information. Social media platforms[2] for investors have recently emerged as popular information-sharing platforms where investors can share ideas, learn investment techniques, and recommend stocks based on their analyses.[3] This increases investors' ability to communicate information. Investor-oriented social media platforms provide an opportunity to observe disagreement among investors, including the heterogeneity of investors and the salience of information signals. Compared to traditional proxies of disagreement such as abnormal trading volume, volatility, and analyst forecast dispersions, investor-oriented social media platforms can offer better insights for the following reasons. First, investor-oriented social media platforms can provide a unique opportunity to observe the heterogeneity[4] of investors. Second, such platforms can provide direct evidence of disagreement arising from investors' social interactions when disclosing their recommendations[5]. Third, the salience[6] of information signals is a decisive factor for investors to allocate their attention efficiently. On such platforms, the salience of information signals can also be observed directly. Following Cookson and Niessner (2019), we use a novel data set from one of the largest investor-oriented social media platforms, StockTwits, to construct a disagreement proxy that is entirely based on social interactions[7] among investors. StockTwits provides useful insights to observe social interactions among investors and the salience of their information signals.

Return synchronicity has been widely used as a proxy for price informativeness.[8] Prior research argues that return synchronicity plays a pivotal role in understanding the extent to

---

[1] Veldkamp (2006) explores the role of information-driven comovement in financial markets and highlights the need for reliable proxies for information acquisition by investors.

[2] Social media platforms are an important byproduct of technological advances. For example, Facebook and Twitter revolutionized the concept of online interactions in the first decade of the 21st century. Similarly, StockTwits started its Twitter-like cashtags service in 2008 for investors, and later Twitter adopted the same technology by adding a cashtags service to its online platform in 2012.

[3] In addition to StockTwits, which is mainly used for brief discussions and ideas by anyone in the investment industry. SeekingAlpha is another prominent investor-oriented platform offering stock market analysis to its users since 2004.

[4] Heterogeneity of agents is a key ingredient of belief dispersion (Kandel & Pearson 1995).

[5] Not all but StockTwits and Seeking Alpha are two investor-oriented platforms where investors can voluntarily disclose their recommendations.

[6] For details see Huang et al. (2018) and (Li *et al.* 2019).

[7] Hong and Stein (1999) highlight the need to develop an interaction-based model that incorporates the follow of information in the financial markets. Moreover, Hong et al. (2004) provide evidence that social interactions among investors increase stock market participation.

[8] For example, Sila et al. (2017) present that the reputation of independent directors is directly linked to increased price informativeness, Mathers et al. (2017) find firms with less synchronicity have better innovation outcomes.

which a stock comoves with industry and market factors. Roll (1988) finds low R^2 statistics in the absence of any news, suggesting that the capitalization of private information through the trading activities of informed arbitrageurs. According to the price informativeness explanation, less return synchronicity reflects more firm-specific information. Therefore, investors will benefit when trading by using that firm-specific information to allocate their resources more efficiently (Morck *et al.* 2000; Durnev *et al.* 2003). Moreover, investors will benefit when they wish to diversify their risk and adjust their portfolio betas (Huang *et al.* 2019a).

Our results show that disagreement among investors on StockTwits decreases return synchronicity. Specifically, one standard deviation increase in disagreement results in a 5.7% decrease in return synchronicity, suggesting an inflow of firm-specific information into the financial markets. These results remain robust after controlling for the effect of media coverage, analyst coverage, and macroeconomic trends, in addition to other firm-level controls. Similarly, these results are robust to controlling firm, time, and industry fixed effects. We also use two-dimensional clustering for firms and months to deal with any serial correlation at the firm level and any systematic shocks over time (Petersen 2009). Our findings suggest that social interactions play a key role in influencing investors' behaviors in financial markets (Hong *et al.* 2004; Hirshleifer 2019).

However, there is mixed evidence on the role of return synchronicity to predict stock price informativeness. For example, Chan et al. (2013) find that liquidity is positively associated with return synchronicity, suggesting low R^2 statistics reflect information asymmetry. Barberis et al. (2005) presented evidence against fundamentals-based views by suggesting that stocks listed in the S&P 500 index comove more with index stocks, and non-listed stocks comove more with non-listed stocks.[9] They argue that irrational investors trade in financial markets based on category view, habitat view, and information-diffusion view. Therefore, frictions or noise could lead to comovement if there are limits to arbitrage[10]. In addition, it is also possible that noise trading increases the magnitude of idiosyncratic pricing errors, which are responsible for low R^2 statistics. Therefore, noise can distort the measure of

[9] Green and Hwang (2009) studied comovement before and after stock splits and present evidence supporting Barberis et al. (2005). They argue that stocks move more with high-priced stocks before splits and low-priced stocks after splits, thus following a category view-based approach.
[10] Based on the frictiona-based and sentiment-based explanations of comovement, disagreement among investors may increase frictions or sentiment, thereby inducing a high R^2 stastistics.

price informativeness and low R^2 statistics do not necessarily indicate deteriorating informational efficiency.

In our research setting, noise is of less concern since StockTwits is an investor-oriented platform. However, to further test whether disagreement on StockTwits provides firm-specific information, it is important to disentangle the difference between noise and price informativeness. To test this conjecture, we implement different tests. First, following Ayers and Freeman (2003), we combine disagreement with leads, contemporaneous, and lag changes in earnings to predict current stock returns. If disagreement among investors reflects noise, disagreement should not increase the price leads of earnings, and the impact should go in the opposite direction for post-earnings-announcement drift. In contrast, we find clear evidence that disagreement increases price leads of earnings and suggest that disagreement accelerates the pricing of future earnings and generates an improvement. Second, we explore individual investors' recommendation revisions on StockTwits. Hong and Stein (2007) provide evidence that suggests that disagreement initially arises due to heterogeneous priors and differential interpretations of information signals by investors in financial markets. Therefore, we assume that investors will revise their recommendations when they update their economic models,[11] based on either their priors or differential interpretation, increasing disagreement among investors and the inflows of firm-specific information in the financial markets. We find consistent results in support of this assumption.

Third, we test to what extent the impact of disagreement on return synchronicity varies across firms with different levels of media coverage.[12] Our findings highlight the two key roles of disagreement among investors on StockTwits. First, StockTwits acts as a catalyst by offering firm-specific information when there is no or low information available from traditional information channels, suggesting disagreement from investor-oriented platforms containing firm-specific information rather than market-wide information. These results are consistent with Roll (1988), who present evidence in favor of low R^2 in the absence of news. Second, StockTwits acts as an intermediary and exacerbates the flows of firm-specific information into financial markets when there is greater media coverage. These results are consistent with Dang

[11] This argument is also in line with Kandel and Pearson (1995) and Hong and Stein (1999) suggesting that investors revise their economic models when their marginal utility of consuming new information is higher than the already available information in the financial markets.
[12] Unlike Dang et al. (2020), in addition to using aggregate media coverage as a control variable, we segregate media coverage based on news types, news topics, and news sources.

et al. (2020) and complement their findings by highlighting the role of social media platforms in the financial markets.

Endogeneity could be a concern in our results as there may be selective coverage of firms on StockTwits, which depends on several exogenous factors as well as biased recommendations due to affiliations with the sample firms. To deal with the endogeneity issue, we first use a two-stage least square (2SLS) instrumental variable approach. We then use two instrumental variables for disagreement. First, we construct a unique proxy of local investors, *Proximity*, which captures the social distances between investors and the firms' headquarters for whom they have been discussing and sharing ideas on StockTwits. Our second instrument is motivated by the role of labor unions in US firms, defined as the total number of issues between labor unions and firms aggregated monthly.

These instruments provide an independent source of exogenous variations for each endogenous regressor and meet the criteria for a valid instrument. First, for the relevance restriction, local investor and labor issues can provide a higher level of firm-specific information. Since StockTwits is an investor-oriented platform, it gives local investors a unique opportunity to share their opinions. Similarly, *Labor_Issues* is positively associated with disagreement as it can exacerbate the number of discussions on social media platforms.

Second, for the exclusion restriction, it is doubtful that price-based comovement can directly affect investors' locations and labor unions. In other words, both instruments should affect only return synchronicity via social media channels as information from these variables is sublimed further by a large network of investors who are experts in their fields. Consequently, such social interactions exacerbate the flow of firm-specific information. Our results, based on the 2SLS approach, force the exogeneous portion of disagreement to explain return synchronicity and leave our main results unchanged. In addition, considering that disagreement on StockTwits is a choice for investors, this can depend on several exogenous factors.[13] Under such circumstances, self-selection bias could be an issue that influence OLS estimates (Heckman, 1979). To address this concern, we implement the two-stage Heckman (1979) selection model. Our results remain unchanged after controlling for self-selection bias.

[13] For example, investors' education, investment type, background, and willingness to participate in different types of communication.

Next, we investigate the role of disagreement in influencing the firm-information environment. The Information environment plays a vital role in determining the influence of agents' learning behaviors in financial markets (Vega 2006). The intuition behind this idea is that if disagreement among investors on StockTwits exacerbate the flows of firm-specific information, they should assist stakeholders by offering firm-specific information to correctly calculate the fair value of the firm (Lin *et al.* 2011), leaving less room for managers to conceal self-serving behaviors (Jin & Myers 2006). Therefore, we would expect the impact of disagreement on return synchronicity to be more pronounced when the firm-information environment is less transparent.

To test this hypothesis, we first use discretionary accruals as a proxy for firm opacity (Hutton *et al.* 2009b). Second, we explore the extent to which the impact of disagreement on return synchronicity varies with firms' diversity. Bushman et al. (2004a) present evidence that the firm's diversity limits the transparency of firms' operations for outside investors. Therefore, we define diversity as the number of business segments and geographic locations the firm operates in and use it as a second proxy for the information environment (Markarian & Parbonetti 2007). Third, we examine differences in the impact of disagreement on return synchronicity across firms with different Industry Concentration levels.[14] Ali et al. (2014) argue that firms in highly concentrated industries disclose less information since the proprietary cost of information is higher in those industries.

Finally, we explore firms that are subject to insider trading activity. Piotroski and Roulstone (2004) present evidence that insider trading activity is a vital private information source for investors. Therefore, we expect that disagreement can facilitate the incorporation of private information into financial markets. Overall, our results show that disagreement has a higher impact on return synchronicity for opaque firms, firms with greater diversity, firms with greater industry concentration, and firms with more insider trades. These findings provide substantial evidence that investors' social media platforms facilitate the incorporation of firm-specific information when firms have a less transparent information environment.

One of the key characteristics of social media is its attention-grabbing features, known as salience (Fiske & Taylor 2013). It is pertinent to note that the scarcity of cognitive resources limits investors' ability to allocate their attention (Kahneman 1973). Therefore, salience plays

[14] We use total assets to calculate industrial concentration based on the Herfindahl–Hirschman index (HHI). However, our results remain consistent when we use total sales to calculate industry concentration.

a pivotal role in guiding investors to allocate their attention. Previous studies have mainly discussed the impact of limited attention without explaining the effects of salience. Therefore, our next strand of investigation is to understand the role of salience. Specifically, we aim to disentangle the impact of disagreement associated with salience. We classify the salience based on information signals and the heterogeneity of investors.

The first salience group, which is based on information signals, is further divided into two subgroups based on their network and social media attention (SMA), where the network is defined as the reach of information signals, and SMA is further divided into the number of ideas on StockTwits, the popularity of those ideas, and discussion threads created by investors on StockTwits. The results show that when investors with large numbers of followers post ideas, other investors follow their lead, thus increasing their influence. Consequently, the higher salience of information signals attracts more audiences to interact with each other, increasing the level of disagreement and prompting higher inflows of firm-specific information into financial markets.

The second salience group, based on investors' heterogeneity, is further divided into three subgroups based on unique investors' presence, investors' self-disclosed investment experience, and approaches. First, our findings suggest that unique investors' arrival increases the impact of disagreement on return synchronicity. These results are consistent with Kandel and Pearson (1995), who argue that heterogeneity of agents is a key ingredient for disagreement among the investors. Second, we calculate within-group disagreement among investors based on their investment experience (professional, intermediate, or novice). Our results show that although professional investors take a lead role in assisting investors by facilitating inflows of firm-specific information, intermediate and novice investors also play a pivotal role. Third, we calculate within-group disagreement among investors based on their investment approaches (momentum, technical, fundamental, or value). Our results suggest that momentum and technical investors provide higher inflows of firm-specific information than investors with fundamental or value investment approaches. These findings are consistent with previous literature highlighting the role of media in increasing momentum investing (Hillert *et al.* 2014).

Our research contributes to the existing literature by offering direct evidence that discussions on investors-oriented social media platforms such as StockTwits provide firm-specific information. Overall, this study contributes to the following areas. First, unlike previous studies that only focus on the role of social media in financial markets to predict volumes and returns, our study provides an essential piece of the puzzle by explaining that a social media platform for investors can predict financial markets because it allows firm-specific information to flow to investors who actively participate in discussions on such platforms and update their priors based on available information. We accomplish this by using disagreement as a unique proxy for discussions among investors on StockTwits and providing substantial evidence that such disagreement can predict stock-return synchronicity. These findings are consistent with Hong and Stein (2007) disagreement model. To the best of our knowledge, this is the first study to provide evidence that chatter on social media platforms assists investors by offering firm-specific information.

Second, this study contributes to the existing literature investigating the firm information environment's role in financial markets. These results provide substantial evidence by highlighting the role of disagreement among investors in a less transparent information environment, thus suggesting that for firms with higher informational opacity, greater diversity, higher industry concentration, and more insider trading, social media platforms for investors act as a catalyst to assist investors by providing firm-specific information.

Third, our study contributes to the emerging literature on the role of the salience of information signals in financial markets. We show that a large social network of investors and a large number of unique ideas posted on StockTwits amplify the impact of disagreement on return synchronicity. Moreover, dissecting investors' heterogeneity on StockTwits further suggests that although geographical background and investors' experience matter, what matters more is the quality of information signals arriving in financial markets and investors' heterogeneity (Kandel & Pearson 1995).

Finally, our study contributes to distinguishing the debate on whether stock prices for companies with a low R^2 statistics is more informative. This study provides substantial evidence using a unique dataset of investor-oriented platforms that less synchronicity can reflect higher stock price informativeness.

Our findings are consistent with the existing literature in providing substantive evidence that social media platforms for investors can predict financial markets. However, no previous study has discussed whether disagreement on social media platforms for investors can provide firm-specific information to the best of our knowledge. Our work builds on these differences and provides further evidence that investors' disagreement on StockTwits provides firm-specific information that can predict financial markets. Our study is closely related to that of Ding *et al.* (2019), who used the number of articles published on SeekingAlpha to predict return synchronicity. Our study provides evidence that, besides attention, social interactions among investors on social media platforms can predict return synchronicity. In relation to these findings, we are the first to provide direct evidence that disagreement on StockTwits provides firm-specific information that can be useful for a broad range of stakeholders, mainly retail investors and portfolio managers in financial markets.

The rest of the paper is organized as follows: Section 2 discusses the literature review and hypothesis development, Section 3 describes the data and explains the research design, Section 4 discusses the empirical results, Section 5 presents robustness checks, and Section 6 presents the conclusion.

2. Literature Review and Hypothesis Development

2.1. Background on Return Synchronicity

Return synchronicity is defined as the extent to which stock return comoves with the market and industry returns (Durnev *et al.* 2003). Therefore, higher return synchronicity means stock returns are explained by industry and market returns, and low synchronicity means the variation in stock returns has a weak association between market and industry returns. Roll (1988) provides corroborative evidence of the weak association between stock returns and industry and market movements. He further suggests that this weak association is attributed to firm-specific information incorporated in the stock prices. Similarly, Shiller (1989) presents further evidence that UK and US firms' dividends cannot fully explain the comovement of stock prices between two countries. To test whether less return synchronicity provides firm-specific information, Durnev *et al.* (2003) provide evidence that firm-specific stock price variability is positively correlated with price informativeness measures. Thus, less synchronicity provides higher firm-specific information.

Building on the exciting research agenda of Roll (1988), further studies have confirmed that various factors determine return synchronicity. In this vein, Pindyck and Rotemberg (1993) highlight the role of market segmentation partly explained by firm size and institutional ownership to describe the individual stock comovement; Morck et al. (2000) present a cross-country sample showing the negative association between return synchronicity and government protection of property rights[15]; Durnev et al. (2004) argue that less synchronicity (higher firm-specific information) is associated with the efficient allocation of capital by the firms; Chan and Hameed (2006) provide evidence that higher analyst coverage is associated with higher return synchronicity, suggesting a lower inflow of firm-specific information by the analysts; Jin and Myers (2006) argue that a lack of transparency increases return synchronicity; and An and Zhang (2013) present a negative association between return synchronicity and crash risk, among others.

Although previous studies provide substantial evidence that less return synchronicity provides firm-specific information, Barberis et al. (2005) provide evidence against this fundamentals-based view. Based on univariate analysis, they argue that additions in the S&P 500 index suggest that a higher level of firm-specific stock variation is synchronized with market movements and has nothing to do with firm-specific information. In similar lines, Green and Hwang (2009), using stock splits, present evidence supporting Barberis et al. (2005), suggesting that the firm-specific information cannot explain price-based comovement. Conversely, Chen et al. (2016) provide corroborative evidence that the inclusion of momentum stocks likely explains some of the sample results reported by Barberis et al. (2005) and Green and Hwang (2009). This is because, after univariate analysis, factor loading based on Dimson (1979), and matching techniques, beta changes are indistinguishable before and after index additions and stock splits. These studies highlight the need to further investigate the relationship between return synchronicity and price informativeness using better identification strategy and datasets.

2.2. Return Synchronicity and Firm Information Environment

The firm information environment plays a pivotal role in determining the flow of firm-specific information in the financial markets. In this vein, Piotroski and Roulstone (2004) argue that, contrary to the existing notion that higher analyst coverage provides firm-specific information, higher analyst coverage exacerbates market information flow, and Jin and Myers (2006) provide evidence that less transparency is associated with higher return synchronicity. Veldkamp (2006) argues that when investors consume information from a common source, information about one asset affects the price of other assets. Therefore, this search for information makes asset prices more efficient and causes some assets to comove with other

[15] This negative association of property rights is also consistent with Roll (1988) since weak enforcement of such rights is likely to impede firm-specific informed trading.

assets. One of the key challenges highlighted by Veldkamp (2006) when seeking to understand information-driven return synchronicity is the lack of data on proxy investors' information.

Recent studies by Ding et al. (2019) using the number of articles on the social media platform argue that social interaction among investors increases the flow of firm-specific information in financial markets. Drake et al. (2017) offer compelling evidence that the comovement of investors' attention has significant consequences on the comovement of returns between the firm and its peers. Jiang et al. (2019) present evidence that co-discussed stocks are actively traded and have higher comovement with the stocks which are discussed together. Similarly, Huang et al. (2019a) provide unique evidence that investors pay more attention to market-level information due to scarce cognitive resources. In the presence of attention shocks, market-level information increases the marginal utility of information consumed by investors.

Although these studies provide compelling evidence that attention plays a pivotal role in determining stock comovement, Shiller (1992) suggests that investing is a social activity in which investors share their opinions about different investment approaches[16]. Moreover, recent studies by Hirshleifer and Teoh (2009) and Hirshleifer (2015) suggest that behavioral and psychological aspects of investors' decision-making are based on current information and priors. Therefore, these studies warrant further evidence to investigate the association between social interactions and firm-specific information in financial markets.

2.3. The Role of Disagreement in Financial Markets

Disagreement arises due to differences in opinion. However, it is pertinent to ask whether disagreement converges or if investors agree to disagree with each other based on the assumption that all investors have heterogeneous priors and the ability to interpret new information in an entirely different fashion (Harris & Raviv 1993). In this research line, Aumann (1976) argues that convergence in disagreement occurs when investors have common priors and a shared understanding of each other's posterior beliefs. However, later studies provide contrasting evidence that disagreement persists in financial markets due to the quality of information signals and investors' uncertainty to interpret them (Varian 1985; Kandel & Pearson 1995; Acemoglu *et al.* 2006).

[16] Hong et al. (2004) provide evidence consistent with Shiller (1992) and argue that social interactions among investors increases the stock market participation.

It is also important to note that the convergence of disagreement depends entirely on the probability of learning from investors' information signals. However, there is ample evidence of investors' inconsistent learning in financial markets (Banerjee *et al.* (2019), e.g., investors' over- and underreactions (Barberis *et al.* 1998) to prices, investors' overconfidence (Odean 1998; Eyster *et al.* 2019), excess volatility, and returns in financial markets. Overall, there is a consensus in the literature that differences in opinion drive financial markets. Disagreement does not converge as investors have heterogeneous priors, and the instantaneous flow of information in financial markets induces investors to continuously update their beliefs (Banerjee & Kremer 2010).

One of the critical challenges in financial markets is to find a suitable proxy for investors' disagreement. In this regard, previous studies have commonly used analyst dispersion (Diether *et al.* 2002), returns (Cen *et al.* 2017), and options (Golez & Goyenko 2019). Similarly, with the emergence of social media platforms for investors, recent studies have also used investors' disagreement on StockTwits (Al-Nasseri & Menla Ali 2018; Giannini *et al.* 2019). A recent study by Cookson and Niessner (2019), using data from StockTwits and measuring investors' disagreement within and across groups (investment philosophies), argues that investors' disagreement arises due to differences in investment approaches and investors' experience. These findings are consistent with Kandel and Pearson (1995) and offer insightful evidence by providing a unique proxy for investors' disagreement using social media platforms for investors.

Based on the existing literature, there is a consensus that investors' disagreement plays a vital role in financial markets.[17] For example, Fama and French (2007) revised their assumption about the distribution of future payoffs and suggested that investors' disagreement affects asset prices. However, current literature does not address the extent to which investors' disagreement on social media platforms for investors provides firm-specific information on financial markets. In this line of research, most studies have mainly focused on predicting volumes and returns based on investors' sentiments on Internet discussion boards (Antweiler & Frank 2006) such as Twitter (Bollen *et al.* 2011; Sprenger *et al.* 2014), SeekingAlpha (Chen *et al.* 2014; Campbell *et al.* 2019), and StockTwits (Renault 2017). This study fills this research gap by investigating

[17] Huang et al. (2019b) concluded that management–investor disagreement can influence management to replace CEOs. In another study, Ayotte (2020) argue that firms have an incentive to exploit disagreement among investors by reducing borrowers' cost of funds and directly affecting the capital structure of the firm.

the role of disagreement on investors' social media platforms and predicting return synchronicity.

2.4. Hypothesis Development

Standard asset-pricing models assume that financial markets can process new information at a sufficient speed to keep pace with information arrivals in financial markets. Consequently, due to the instantaneous flows of information, asset prices adjust according to the available information. This assumption is consistent with rational expectation models in frictionless markets. However, in a market with friction and limits on arbitrage, information friction, endowed biases among investors and various stakeholders in financial markets, and scarce cognitive resources challenge investors' ability to process public and private signals.[18] A substantial body of literature focuses on understanding how investors process information in financial markets. For example, Daniel et al. (1998) discuss the role of psychological factors to explain the under- and overreactions of investors to specific information signals, belief heterogeneity, and differences of opinion (Banerjee *et al.* (2009).

The disagreement hypothesis is based on three main assumptions explained by Hong and Stein (2007) disagreement model. The first is the flows of information in financial markets, suggesting that even in the presence of a continuous stream of information, not all investors consume information at the same level due to scarce cognitive resources. The second is that investors have limited attention due to scarce cognitive resources. This assumption suggests that information released in an attention-grabbing[19] manner will have more implications for investors than general news flowing into the market. The third assumption suggests that investors have heterogeneous priors, and while all investors may receive the same information simultaneously, they interpret that information based on their heterogeneous priors. The disagreement model offers a natural framework for understanding disagreement in financial markets, explaining mechanisms that can generate investor disagreement. In line with Hong and Stein's (2007) disagreement model, we argue that, due to the instantaneous flows of information in financial markets via different information channels, investors on social media platforms such as StockTwits consume information from every available information channel

[18] Eyster et al. (2019) propose a cursed trader's model based on private information acquired by other investors in financial markets. They conclude that cursed traders actively trade and assume excessive risk in financial markets. Vives and Yang (2017) propose a model of costly information, whereby investors' interpretation of prices is noisy due to bounded rationality.
[19] Attention-grabbing features of information signals include categories such as the salience of information signals.

and participate in discussions on StockTwits using cashtags. Those investors interpret information signals based on their heterogeneous priors[20] and participate in discussions by sharing their ideas and recommendations. Since investors will only update their recommendations based on their interpretation of information signals on StockTwits, frequent updates to investors' recommendations will increase disagreement between investors. Increased disagreement among those investors offers a unique opportunity to understand the information acquisition process in financial markets and the social benefits resulting from the interactions between investors on such platforms. In line with these arguments, we expect that disagreement on StockTwits adds additional value to investors' economic models by offering firm-specific information. This leads to our main hypothesis:

H1: *Ceteris paribus, greater disagreement among investors on StockTwits results in a decrease in return synchronicity.*

3. Data Description and Research Design

3.1. Data Description

Our firm-level data come from NYSE/AMEX/NASDAQ-listed companies' common stocks with share codes of 10 and 11 from January 2013 to December 2017. Stock data are collected from the Centre for Research in Stock Prices (*CRSP*). Quarterly firm-level financial statements data are collected from Compustat, and I/B/E/S and insider-trading data are collected from Thomson Reuters. We download firm-level media-coverage data from Ravenpack News Analytics (RPNA) services. RPNA offers greater flexibility and in-depth media coverage as compared to any other news database. Overall, we download 2.2 million news articles with daily coverage of our sample firms. The daily news articles are aggregated monthly to calculate media coverage. To ensure that we only download relevant news data, we apply RPNA *Relevance* and *Novelty* filters.[21] The average *Relevance* score for our media coverage data is 89.90, with a median of 98 and a standard deviation of 15.45. Similarly, the average *Novelty* score is 94.03, with a median of 100 and a standard deviation of 12.42. Overall, the average number of news articles is 2.95, with a median of 2.94 and a standard deviation of

[20] Harris and Raviv (1993) and Kandel and Pearson (1995) call these heterogenous priors economic models.

[21] Ravenpack news analytics standardizes Relevance and Novelty scores from 0–100. In the case of relevance, a score of 0–100 indicates how strongly a firm is related to a news story; values greater than 50 suggest that the firm is discussed along with other firms, and values greater than 75 suggest the firm is discussed significantly in the news. In the case of novelty, a score of 0–100 indicates how novel the news story is within a given 24-hour period. Values greater than 50 suggest that multiple stories exist within the 24-hour period, and values equal to 100 suggest that the news story is completely novel within the same period.

1.06 articles per month. To further ensure data relevance and exclude small and less traded firms, we only include firms with an average share price above $1, and whose trading data for the last year are available in CRSP. We then match our sample firms with the StockTwits sample. Based on data availability while matching with different databases, we have 956 firms in our final sample with 53,778 firm-month observations from 2013 to 2017. The overall average *Firm Size* is $19.33 billion with a median of $3.21 billion, and the average *Analyst Coverage* is 12 analysts for each firm.

To construct disagreement as our main variable of interest, we collect data from StockTwits, a popular social media platform for investors. StockTwits is by far the largest social media community for investors and traders, with more than 3 million members, 5 million monthly messages, and 3 million monthly visitors.[22] The main user interface of StockTwits is user-friendly; investors can post Twitter-like messages up to a 140-character limit.[23] One of the distinguishing features of StockTwits is investors' user profiles, where any investor can volunteer to disclose their asset choices, investment approaches, and investment term preferences. Moreover, investors can create a customized watchlist to view StockTwits' ideas directly relevant to their investment preferences. To collect data from StockTwits, we developed a python program to connect with StockTwits API and collect multiple data points based on numerous iterations from January 2012 to December 2018. StockTwits relies on cashtags ($AAPL) as company identifiers. We were able to harvest more than 38 million ideas during our data collection, posted by approximately 297,000 users, discussing 8,394 companies listed in the US and other national stock exchanges.

We apply several filters to ensure the quality of data harvested from StockTwits. To this end, to ensure that data are relevant, we only keep data from January 2013 to December 2017. To avoid concerns related to shared attention and discussion of multiple cashtags in a single message, we only keep StockTwits ideas that contain single cashtags and only discuss one company in a given message. Our final filter restricts our StockTwits sample to those companies listed in NYSE/AMEX/NASDAQ as common stocks. This filter is essential in terms of investors' ability to identify companies and discuss their ideas uniquely. After applying these filters, we are left with approximately 16 million ideas containing 1,890 unique cashtags.

[22] https://about.stocktwits.com/

[23] On May 8, 2019, StockTwits increased their character limit to 1000 characters. However, this occurred outside our sample period.

We match the StockTwits sample with our US firm-level data, resulting in 956 firms and 53,778 firm-month observations in our final sample, with more than 12 million ideas.

[Insert Table 1 Here]

3.2. StockTwits: Where Is It All Coming From?

StockTwits is a social media platform for investors. With the recent surge in information flows in financial markets, it is crucial to understand that ideas expressed on StockTwits must be accurate representations of investors' opinions. We use StockTwits for the following reasons. First, the heterogeneity of investors is a distinct attribute of StockTwits. Investors can voluntarily disclose their experience level (professional, intermediate, or novice) and can disclose their investment approaches. Second, StockTwits is free for anyone wishing to share and post their ideas. This motivates investors to engage in discussions with other investors. Third, during this information-sharing process, investors have the opportunity to increase their social influence by increasing their number of followers after sharing unique investment ideas and expert analysis. Therefore, StockTwits encourages investors to share their ideas and learn from investment professionals who are experts in their fields.

It is also vital to understand the dynamic structure of StockTwits compared to other social media platforms for investors. In a recent study, Cookson and Niessner (2019) argue that investors on StockTwits are less inclined to post fake news and more interested in posting reliable information to become famous. Clarke et al. (2020), using data from SeekingAlpha, differentiate between fake news and legitimate news articles. They argue that although fake news articles attract investors' attention, legitimate news articles generate higher trading volumes. However, it is pertinent to note that investors' motivation to post on StockTwits as compared to SeekingAlpha is entirely different. On SeekingAlpha,[24] authors are paid for their articles if they start to attract a certain number of readers; on StockTwits, users post without seeking any monetary benefit.[25] Therefore, unlike any other social media platform, ideas posted on StockTwits are less subject to potential bias.

[24] SeekingAlpha payment terms and conditions can be found here: https://seekingalpha.com/page/payment-terms

[25] On April 10, 2017, the SEC cracked down on alleged stock-promotion schemes on SeekingAlpha, whereby some authors were paid to write in favor of certain companies listed in the United States. Further details can be found here: https://seekingalpha.com/article/4061813-seeking-alpha-applauds-secs-actions-to-stomp-out-stock-promotion

Next, misinformation does not seem to be a matter of concern in our data set for the following reasons. First, it is highly unlikely that investors on StockTwits can influence financial markets by sharing false information. This is because several other information channels will have a crowding-out effect on these investors' incorrect information. Second, investors' primary motivation to share opinions on such platforms is to gain popularity on them, and for that reason, disseminating false information and fake news can be harmful to their StockTwits profile. Third, there are no financial benefits for investors who post their ideas on StockTwits, except sharing ideas and recommending stocks. More importantly, unlike Twitter, StockTwits is not seen as a marketing platform for individuals. Finally, since our sample firms have large market capitalization and high liquidity, it is less likely that misinformation (if any) on StockTwits can influence these stocks' prices.

Our StockTwits sample contains more than 12 million ideas,[26] posted by 162,836 distinct investors[27]. Panel A of Table 1 presents summary statistics for the StockTwits data on our sample firms. All StockTwits variables are log-transformed, except disagreement. Overall, the average disagreement on StockTwits is 0.53. The average monthly frequency of ideas on StockTwits is 3.99, posted by 3.75 investors on average, with an average social media experience[28] of more than 13 months. To better understand the impact of an extensive social network, we constructed *Network* as a monthly variable by aggregating the number of followers of distinct investors who post ideas on StockTwits related to specific firms. The average monthly *Network* level is 12.97, with a median of 13.28. We also calculate the number of revisions a distinct investor makes on the same stock. *Revisions* are defined as the sum of the number of times per day a distinct investor revises their recommendations (e.g., Bullish to Bearish or vice versa), which is then aggregated at a monthly frequency. *Revisions* are log-transformed; the average number of *Revisions* is 3.61, and the median is 3.22.

To share ideas on StockTwits, investors must create a user profile. Fig. 1 presents a summary of information from investors' profiles on StockTwits. Panel A of Fig. 1 presents the distribution of investors according to their StockTwits joining year. The x-axis is the investors' joining year, and the y-axis is the percentage of investors joining StockTwits in that specific year. It is pertinent to mention that StockTwits has become more popular in recent years. Panel

[26] StockTwits messages posted by investors are referred to as ideas on StockTwits.
[27] We recognize all users who post ideas on StockTwits as investors.
[28] The average social media experience is calculated as the number of months between the date of joining StockTwits until the investors posted their first idea in our sample.

B shows the distribution of ideas and investors in the sample years, where the x-axis is the sample year. The y-axis is the percentage of investors/ideas on StockTwits discussing the sample firms. It is important to understand the distribution of StockTwits ideas and investors at the sector level since some industries on social media receive more coverage than others. In this regard, Panel C presents the distribution of ideas and investors across different sectors based on the Global Industrial Classification System (GICS) four-digit sectors. It is evident that the healthcare sector has the highest concentration of investors, and consumer discretionary and information technology have the highest concentration of ideas, respectively.

[Insert Figure 1 Here]

Due to the breadth of information in our data harvested from StockTwits, we try to understand the geographical distribution of ideas and investors in the US at the state level.[29] To do this, we standardize users' locations at the city and region levels so that we can allocate ideas and investors to relevant US states using state-level coordinates obtained from the 2017 US Census website geographic data. Fig. 2 presents the distribution of ideas and investors across US states. Their geographical distribution is important for the following reasons. First, it shows that most of the investors who frequently post on StockTwits are based in the USA, with only 2% of investors and ideas coming from the rest of the world.[30] Second, the geographical distribution of investors suggests that the majority of investors come from the three largest states, i.e., California (16.18%), New York (16.17%), and Texas (8.32%).

[Insert Figure 2 Here]

3.3. Variables Construction

3.3.1. Return synchronicity

The measure of return synchronicity is calculated based on the value of the coefficient of determination (R^2). To derive the value of R^2, we use Carhart (1997) four-factor model, since the momentum factor, in addition to the Fama–French factors, which may also be a source of variation in return synchronicity, can better account for idiosyncratic risk. We use the following equation to calculate the value of R^2:

[29] Sixty-two percent of investors out of 162,836 disclosed their location in their StockTwits public profiles. We also lose some investor-level location data in the data-cleaning and standardization process. For example, we cannot map investors who only disclosed their country as their location.

[30] As compared to country-level information disclosed by investors on their StockTwits public profiles.

$$Ret_{i.d} = \beta_0 + \beta_{mkt.i} MKT_d + \beta_{HML.i} HML_d + \beta_{SMB.i} SMB_d + \beta_{UMD.i} UMD_d + \varepsilon_{i,d} \quad (1)$$

Where $Ret_{i.d}$ is the daily return on stock *i* at day *d*. The right side of the equation is market (MKT), high minus low (HML), small minus big (SMB), and momentum (UMD) factors. To ensure the availability of sufficient numbers of daily observations to calculate monthly R^2, the stock return must have at least 50% of non-missing observations on trading days in a given month. $R^2_{i,t}$ derived from Eq. (1) is the coefficient of determination, ranging between 0 and 1 for stock *i* during month *t*. However, the existing measure has a high level of skewness and kurtosis, resulting in some econometric issues. To deal with this problem, we take the natural logarithm of $R^2_{i,t}$, which is consistent with the existing literature (Morck *et al.* 2000; Piotroski & Roulstone 2004). Our final measure presents an unbiased proxy for monthly return synchronicity $Sync_{i,t}$ for stock *i* during month *t* and is calculated as follows:

$$Sync_{i,t} = \ln\left(\frac{R^2_{i,t}}{1 - R^2_{i,t}}\right) \quad (2)$$

Summary statistics for $Sync_{i,t}$ are presented in Panel C of Table 1. The average *Return Synchronicity* is –0.37, with a standard deviation of 1.04 and a median of –0.33. To further check the robustness of our synchronicity measure, we also calculate the return synchronicity following Roll (1988), who argues that market and industry returns, as well as a firm's stock returns, are inversely related to the firm-specific information incorporated in stock prices. Following Peng and Xiong (2006) and Anton and Polk (2014), we also use Pearson's correlation coefficient to measure return synchronicity, which is the correlation between firm return and market return. We also use their method to calculate return synchronicity.

3.3.2. Recommendation classification model

Investors' recommendations on StockTwits represent voluntary disclosures. Therefore, not all ideas posted on StockTwits contain investors' recommendations. To construct our measure of disagreement, we use supervised machine-learning classification models to predict investors' recommendations. To this end, a key requirement is to build a robust training data set that is sufficiently large and accurate. Unlike previous studies that mainly rely on hand-classified training data of up to 3000 posts, we use more than 1 million pre-classified ideas labeled by investors on StockTwits. Therefore, our training data set is both large and accurate. We use Baziotis et al.'s (2017) Ekphrasis library for data pre-processing, which is a specialized text pre-processing[31] tool for online social networking platforms. There is no rule of thumb for choosing classification models. It mostly depends on the type of data set and prediction accuracy after cross-validation tests. We use the Random Forest Decision Trees (RFDT) method to more quickly implement and more easily interpret, based on decision trees (compared to the Support Vector Machine). For example, Fernández-Delgado et al. (2014) tested 179 classifiers and concluded that the Random Forest model is one of the best classification models with a close match for Support Vector Machine (SVM) models. Therefore, in our study, we use the RFDT model for recommendation classification. However, to further ensure our results' comparability and robustness, we also use the SVM and Maximum Entropy (MaxEnt) models.[32]

The recommendation classification process is completed in two steps. In the first step, we create multi-way decision trees from the data set such that the data set is split into smaller subsets to predict target values. To maximize information gain and reduce the level of uncertainty in predictions, we use entropy as our impurity criterion. During the prediction process, conditions are presented as nodes, and possible outcomes are presented as edges. The decision-trees method is advantageous because of its quick application and faster turnover on large training data sets. However, one of the main drawbacks of using decision trees is overfitting. Although tree depth is vital to address the overfitting problem, we use the Random Forest[33] (RF) model in the second step. The Random Forest model operates as an ensemble and uses these decision trees to pool all classifications and predict final recommendations. Our final classification results are based on feature selection and are robust under tenfold cross-validation, with an F1 score of 89% and an overall accuracy of 81%.

3.3.3. Disagreement

To construct our disagreement measure, we follow Antweiler and Frank (2004) approach to calculate the average recommendations as follows:

$$Rec_{i,t} = \frac{Rec_{i,t}^{Bulli} \; - \; Rec_{i,t}^{Bearish}}{Rec_{i,t}^{Bullish} + \; Rec_{i,t}^{Beari}} \in [-1, 1] \tag{3}$$

[31] Further details on data pre-processing are presented in Appendix B.
[32] Further discussion is presented in Section 6, along with regression results based on Eq. (5).
[33] Model derivation is presented in Appendix B.

Where $Rec_{i,t}$ is the average recommendation for firm *i* in month *t*. Similarly, $Rec_{i,t}^{Bullish}$ is the aggregate bullish recommendations, and $Rec_{i,t}^{Bearish}$ is the aggregate bearish recommendations. Our average recommendations range between –1 and 1. Our disagreement measure deviates from Antweiler and Frank (2004) assumption, as they assume latent disagreement when there are no posts. In this regard, our approach is consistent with Cookson and Niessner (2019), who assume that no posting means no disagreement. For this purpose, we normalize no-posting cases equal to 0. Therefore, we calculate the overall disagreement as follows:

$$Disagreement_{i,t} = \sqrt{1 - Rec_{i,t}^2} \in [0, 1] \tag{4}$$

$Disagreement_{i,t}$ in Eq. (4) is the overall disagreement among investors' recommendations for firm *i* in month *t*. The value of $Disagreement_{i,t}$ ranges between 0 and 1, where 0 represents complete agreement and 1 represents complete disagreement between investors on StockTwits. Panel A of Table 1 presents overall disagreement with an average of 0.53, a standard deviation of 0.56, and a median of 0.59.

3.4. Research Design

To examine the relationship between Return Synchronicity and *Disagreement*, we estimate the following model:

$$\begin{aligned} Sync_{i,t} = {} & \beta_0 + \beta_1 Disagreement_{i,t} + \beta_2 Media\ Coverage_{i,t} + \beta_3 Analyst\ Coverage_{i,t} \\ & + \beta_4 Leverage_{i,t} + \beta_5 Adv/Sales_{i,t} + \beta_6 \,{}^{Market}\!/_{Book}\, Ratio \\ & + \beta_7 Firm\ Size_{i,t} + \beta_8 ROA_{i,t} + \beta_9\ Earnings\ Volatility_{i,t} \\ & + \beta_{10} Sales\ Growth_{i,t} + \beta_{11} Real\ GDP_{m-1} + V_i + V_t + V_p + \varepsilon_{i,t} \end{aligned} \tag{5}$$

Where $Sync_{i,t}$ is the *Return Synchronicity* of firm *i* at time *t*. The key explanatory variable in Eq. (5) is the $Disagreement_{i,t}$ among investors on StockTwits about firm *i* at time *t*. We estimate the equation using the fixed-effects estimator to account for unobserved firm-specific heterogeneity (V_i). We also control for time (V_t) and industry (V_p) fixed effects by including month and industry dummies, capturing time-varying and industry-specific movements.

Following Petersen (2009), we use two-dimensional clustering at firm and time levels to account for any within-group correlation that may influence standard errors.[34]

We also employ firm-level and market-level control variables, which may directly or indirectly affect our variables of interest. For example, we use *Media Coverage* to control the effect of firm-specific information from alternative sources of information. Following Chan and Hameed (2006), to control for the effect of analysts following the firm, we use *Analyst Coverage*. Firm Size is used to control the firm's size to attract investors' attention, and because the demand for analysts' recommendations is directly proportional to the firm's size. To account for the creditors' monitoring role, which may influence the firm-information environment, we use *Leverage*. Firth et al. (2008) argue that firms with higher leverage have higher monitoring by creditors, consequently playing a vital role in managers' decision-making. We use firms' *Adv/Sales*[35] and *Market/Book* ratios to control the effect of firm influence on the external information environment and control the effect of firm-level growth opportunities based on the market value of equity, respectively. Similarly, we use *ROA*, *Earnings Volatility,* and *Sales Growth* as indirect proxies for investors' attention in financial markets. Finally, to control for the effect of macroeconomic trends, we use the monthly lagged value of *Real GDP* (Brockman *et al.* 2010). To make the statistics intuitive, we standardize all the right-hand-side variables in our regression. Variable definitions are presented in Appendix B.

4. Empirical Results

4.1. Correlation Table

To understand the relationship between explanatory variables and *Return Synchronicity*, we measure Pearson's correlation coefficient between all variables.[36] The correlation between *Return Synchronicity* and *Disagreement* is significantly negative, consistent with our hypothesis, according to which disagreement might reduce return synchronicity. *Analyst coverage* is an important source of industry and market-level information, and it has a 23.2 and 17.7% correlation with *Return Synchronicity* and *Disagreement,* respectively. *Media Coverage* is another variable of interest since it plays a pivotal role in influencing firms' information environment. It has a 7.4% correlation with *Return Synchronicity* and a 17.5% correlation with

[34] Our results remain consistent when clustering at the firm level only.
[35] Chemmanur and Yan (2019) present that an increase in advertising increases the firm's visibility among investors and attracts investors' attention. In another study, Grullon et al. (2004) present that firms with higher advertising costs have higher liquidity and a large investor base.
[36] For brevity reasons, the table is presented in the online Appendix.

Disagreement. Overall, the correlation between *Return Synchronicity* and the remaining explanatory variables is less than 15%, suggesting no multicollinearity problem in our regressions.

4.2. Main Results

The results presented in Table 2 are estimated using Eq. (5). We use ordinary least square regression (OLS) with the firm, month, and industry fixed-effects as our main regression model. To understand the impact of *Disagreement* on *Return Synchronicity,* Model (1) shows the impact of *Disagreement* without controlling for the effects of firm-level covariates and macroeconomic trends. The result supports our hypothesis that *Disagreement* reduces *Return Synchronicity.* Considering that we standardized the right-hand-side variables, a standard deviation increase of one in *Disagreement* results in 6.9% greater inflows of firm-specific information into financial markets, which is economically significant. One of the key factors that may influence firm-specific information flows in financial markets is *Media Coverage.*[37] Therefore, in Model (2), we add *Media Coverage* to account for the impact of firm-specific news in our model. Our results show that there is a negative relationship between *Media Coverage* and *Return Synchronicity,* suggesting that higher media coverage results in greater inflows of firm-specific information into financial markets.[38] Similarly, the *Disagreement* coefficient remains significant and negative, suggesting that investors actively use social media platforms such as StockTwits and consume information from such platforms in addition to conventional media coverage. The magnitude of the impact of *Disagreement* on *Return Synchronicity* in Model (2) suggests that a standard deviation increase of one in *Disagreement* results in 5.4% higher inflows of firm-specific information in financial markets.

[Insert Table 2 Here]

In Model (3), we add firm-level covariates and macroeconomic trends to control for the effects of firm-level factors and macroeconomic trends influencing firm-specific information in financial markets. Specifically, we add *Leverage* and *Analyst Coverage* to control for the effect of the firm-information environment. The results from *Leverage* are consistent with Armstrong et al. (2010), suggesting that higher *Leverage* will increase the monitoring benefits for shareholders. The results for *Analyst Coverage* suggest that greater analyst coverage means

[37] To account for the aggregate effect of media coverage. Later in our analysis, we present the impact of media coverage after dissecting it based on news types, news topics, and news sources.

[38] These findings are consistent with the recent study by Dang et al. (2020), who suggest that firm-level media coverage provides firm-specific information to investors in financial markets.

more focus from analysts on the mapping of industry- and market-level information instead of firm-specific information. These results are consistent with Piotroski and Roulstone (2004) and Chan and Hameed (2006), who concluded that *Return Synchronicity* is positively associated with *Analyst Coverage*. We use the firms' *Adv/Sales* ratio to control the firm's flows of firm-specific information to increase the firm's visibility and attract individual and institutional investors. We use *Firm Size* to control the size of firms to influence the firm-information environment. Da et al. (2011) suggest that several indirect proxies of attention may affect information flows in financial markets. To account for such factors, we use firm-level earnings (*ROA*), *Earnings Volatility,* and *Sales Growth* as indirect proxies for investor attention, which may influence the firm-information environment. Following Chue et al. (2019) and Brockman et al. (2010), we add the lagged value of $Real\_GDP_{t-1}$ to Model (3) to account for macroeconomic trends and their implications for *Return Synchronicity*.

Our results in Model (3) remain consistent after controlling for the effects of firm-level covariates and macroeconomic trends, suggesting that a standard deviation increase of one in *Disagreement* results in 5.7% higher inflows of firm-specific information in financial markets. To further test the validity of our results, we run OLS regression in Model (4) without using any fixed effects. Our results remain consistent in Model (4). Similarly, to account for any cross-sectional correlation and estimate consistent standard errors, we use Fama and MacBeth (1973) two-step regression. The results presented in Model (5) remain consistent with our previous findings.

4.3. Evidence of Firm-Specific information from Disagreement

In this section, we explicitly test our assumption that disagreement on StockTwits provides firm-specific information to investors in financial markets. For this purpose, we use different identification strategies (*Price Informativeness*, *Recommendation Revisions,* and *Media Coverage*) to explore the role of disagreement in predicting firm-specific information.

4.3.1. Price Informativeness

According to Roll (1988), a lower value of R^2 is due to arbitrageurs who gather and possess private information while trading in financial markets. This is because he could not find any association between firm-specific[39] stock-price movements and news releases, and he further

[39] Durnev et al. (2003) test this hypothesis and provide further evidence that greater firm-specific price risk is associated with higher stock-price informativeness. This is consistent with Roll (1988) argument that firm-specific information reflects the presence of arbitrageurs capitalizing based on private information in financial markets.

suggests that "The financial press misses a great deal of relevant information generated privately" [p. 564]. In our setting, we argue that investors' social media platforms offer a unique opportunity for investors to share their ideas. Under this assumption, if investors on StockTwits are discussing firm-specific information, the incorporation of price informativeness should illustrate the impact of disagreement to predict future earnings. We follow Ayers and Freeman (2003) model to test whether *Disagreement* on StockTwits improves the prediction accuracy of future earnings of firms and estimate the following regression equation:

$$\begin{aligned} CAR_{i,t} = {} & \beta_0 + \beta_1 Disagreement + \beta_2 \Delta Earnings_{t+1} + \beta_3 Disagreement \\ & * \Delta Earnings_{t+1} + \beta_4 \Delta Earnings_t + \beta_5 Disagreement * \Delta Earnings_t \\ & + \beta_6 \Delta Earnings_{t-1} + \beta_7 Disagreement * \Delta Earnings_{t-1} \\ & + \beta_8 Controls + \varepsilon_{i,t} \end{aligned} \tag{6}$$

All variables are calculated at the quarterly frequency to maintain consistency between *Disagreement* and earnings variables.[40] $\Delta Earnings$ is calculated as the change in earnings of firm *i* from quarter *t-1* to *t* scaled by the market value of the equity of firm *i* at the beginning of the quarter.[41] $\Delta Earnings_{t+1}$, $\Delta Earnings_t$, and $\Delta Earnings_{t-1}$ are proxies for lead, contemporaneous, and lag changes in earnings, respectively.

$CAR_{i,t}$ is the cumulative abnormal return of firm *i* calculated as the sum of daily abnormal returns for quarter *t.* Abnormal return is calculated as follows:

$$AR_{i,t} = R_{i,t} - R_{SBM,t}$$

Where $AR_{i,t}$ is the abnormal return for firm *i* in each quarter *t*, $R_{i,t}$ is the daily return for firm *i* and $R_{SBM,t}$ is the equal weighted 5 × 5 portfolio return calculated using size and Book/ Market value of equity *(BE/ME)* of firm *i*.

Our main variables of interest are the interactions between *Disagreement* and $\Delta Earnings_{t+1}$ and between *Disagreement* and $\Delta Earnings_{t-1.}$ The intuition is that if low return synchronicity implies more firm-specific information is reflected in prices, we would expect that increased *Disagreement* can increase price leads of earnings. If disagreement among investors reflects noise, disagreement should go in the opposite direction for post-

[40] Ayers and Freeman (2003) use annual data instead of quarterly data. However, social media platforms for investors post ideas based on a short-term approach. In this regard, Giannini et al. (2019) present evidence that investors' opinions either converge or diverge around earnings announcements, without having any implications in the long run.

[41] We use Compustat data item 18 (income before extraordinary items) as a measure of earnings.

earnings announcement drift (*PEAD*). We use the impact of $\Delta Earnings_{t+1}$ and $\Delta Earnings_{t-1}$ on abnormal returns to identify price leads of earnings and the post-earnings announcement drift.

[Insert Table 3 Here]

Table 3 presents the regression results estimated using Eq. (6). We use OLS regression with firm-level fixed effects as our main model and then test the robustness of our results using simple OLS and Fama–MacBeth regressions, respectively. Model (1) presents the stand-alone regression results where only contemporaneous earnings are positive and statistically significant at the 1% level, suggesting that security price responds to contemporaneous earning rather than lead and lag changes in earnings. Model (2) presents the regression results for the interaction between *Disagreement* and change in earnings at the leading, contemporaneous, and lagged levels. If the increase in price informativeness illustrates the impact of *Disagreement* on incorporate future earnings, then β_3 should be positive. Our regression results in Model (2) present that coefficient β_3 is positive and statistically significant at the 1% level, which is consistent with our projection that security prices reflect future earnings when there is more *Disagreement* among investors. In contrast, the coefficient β_7 is negative and statistically significant at the 1% level, suggesting that disagreement among investors does not reflect noise. *Disagreement,* in fact, reduces the price response to $\Delta Earnings_{t-1}$. These results are in line with our hypothesis that *Disagreement* increases price informativeness; firms with high levels of *Disagreement* have lower post-earnings announcement drift (*PEAD*).

Next, the interaction between $\Delta Earnings_t$ and *Disagreement* is negative but insignificant, suggesting that *Disagreement* has no impact on price responses to contemporaneous earnings. It is pertinent to note that the insignificant interaction between contemporaneous earnings and *Disagreement* implies that the lead effect of *Disagreement* is not due to changes in the magnitude of price responses to earnings. If firms with more *Disagreement* have higher earnings persistence than others, we should also find a positive association between contemporaneous earnings and *Disagreement.* In our case, the coefficient from β_5 is insignificant, thus further strengthening our results from β_3. Furthermore, we include firm size as a control in Model (3). Our results remain consistent when we add the control variable.

Models (4) to (6) present the regression results without using firm-level fixed effects. Finally, in Models (7) to (9), we present our results using Fama–MacBeth regression, as in Ayers and Freeman (2003). Overall, our results remain consistent, and the coefficient β_3 is positive. These findings suggest that price leads increase as *Disagreement* increases. Therefore, *Disagreement* on StockTwits is an important source of firm-specific information. Our paper again offers a direct test to distinguish between the two broad theories of comovement (the price informativeness explanation vs. the price noise explanation). Therefore, we provide substantial evidence based on disagreement among investors from an investor-oriented platform and support firm-specific price movements in which less synchronicity can reflect higher stock price informativeness.

One of the key elements of the disagreement model, as explained by Hong and Stein (2007), is heterogeneous priors; i.e., even if the information is released to all investors simultaneously, each investor will interpret information signals based on their economic model. [42] This assumption is important in our context to understand if investors update their beliefs based on their heterogeneous priors and the arrival of information in financial markets. If so, how does the process of belief formation affect the level of *Disagreement* in financial markets?

To understand this mechanism, we use investors' recommendation revisions on StockTwits. A recommendation revision [43] is defined as the number of times distinct investors on StockTwits update their recommendations from Bullish to Bearish and vice versa for a given stock at time *t* compared to *t-1*. The main intuition behind this idea is that investors will only update their recommendations when they update their economic models based on their priors, thus increasing *Disagreement* among investors and the inflows of firm-specific information in financial markets. We estimate the following equation to understand this mechanism:

$$\begin{aligned} Sync_{i,t} = {} & \beta_0 + \beta_1 Disagreement_{i,t} + \beta_2 Revisions_t + \beta_3\, Disagreement_{i,t} \\ & * Revisions_t + \beta_4 Revisions_{t-1} + \beta_5 Disagreement_{i,t} \\ & * Revisions_{t-1} + \beta_6 Controls + V_i + V_t + V_p + \varepsilon_{i,t} \end{aligned} \tag{7}$$

Where $Revisions_t$ is the total number of recommendation revisions for a given stock at time *t,* and $Revisions_{t-1}$ is the lagged value of recommendation revisions for a given stock at time *t-1*. Recommendation revisions are calculated daily and then aggregated monthly. We

[42] Harris and Raviv (1993) and Kandel and Pearson (1995) suggest that differential interpretation of the same signals occurs since investors have different economic models.
[43] It is pertinent to note that Disagreement and Recommendation Revisions are different variables of interest and provide different sets of information. Disagreement is calculated monthly and measures overall disagreement among investors. Recommendation revisions are calculated at the investor level and the number of updates they have made from *t-1* to *t*.

include $Revisions_{t-1}$ in Eq. (7) to understand the impact of stale economic models, i.e., the number of revisions in the last month, on *Return Synchronicity.*

[Insert Table 4 Here]

Table 4 presents the regression results estimated based on Eq. (7). In Model (1), the coefficient β_2 is negative and significant at the 1% level, suggesting that recommendation revisions provide firm-specific information. The coefficient β_4 is also negative and statistically significant. However, the magnitude is significantly smaller as compared to the coefficient of β_2. These findings suggest that stale economic models can predict the less firm-specific information in financial markets compared to the contemporaneous $Revisions_t$. We conduct a *t*-test on the equality of coefficients of $Revisions_t$ and $Revisions_{t-1}$ and reject the null hypothesis that the coefficients are equal. In Models (2) and (3), we examine the interaction between $Revisions_t$ and $Revisions_{t-1}$ with *Disagreement*, respectively, to understand the moderating effect of recommendation revisions on *Return Synchronicity*. As expected, the coefficient of the interaction between *Disagreement* and $Revisions_t$ in Model (2) is negative and statistically significant at the 10% level (–2.5%). These results show that a change in $Revisions_t$ increases the impact of *Disagreement* on *Return Synchronicity,* suggesting that when investors update their economic models, *Disagreement* increases among investors on StockTwits and, consequently, there are higher inflows of firm-specific information in financial markets. The results in Model (3) suggest that $Revisions_{t-1}$ do not affect the impact of *Disagreement* on *Return Synchronicity* since the coefficient of the interaction is not significant. This result also offers useful insights for our analysis by suggesting that *Disagreement* among investors mainly stems from the most recent information.

These findings are consistent with the existing literature. For example, Hong and Stein (2007) argue that investors interpret information based on their economic models even if each investor receives the same information signals, and Banerjee and Kremer (2010) argue that *Disagreement* does not converge since investors update their beliefs due to instantaneous flows of information in financial markets. Therefore, this study provides unique evidence from social media platforms for investors by suggesting how investors on StockTwits updating their beliefs and interactions offers firm-specific information.

4.3.2. Media Coverage

Media coverage is an important source of firm-specific information in financial markets. Previous studies have used the aggregated impact of media coverage. For example, Dyck et al. (2008) and Dyck et al. (2013) highlight the role of media coverage in influencing firms' corporate governance and working in favor of special interest groups, respectively. Fang and Peress (2009) argue that media coverage plays a vital role in influencing stock returns in financial markets, and Chahine et al. (2015) present evidence of strategic communication between managers and the media and highlight the role of informative news in financial markets. In our context, the role of *Media Coverage* is twofold. First, higher *Media Coverage* may attract investors to search for further information about the firm and share their opinions/analyses on StockTwits. Second, investors on StockTwits consume any relevant firm-specific information from other channels of information to create discussion threads on StockTwits, as well as updating their recommendations.

In this study, although we are using *Media Coverage* to control the flows of information from traditional information channels, the conspicuous nature of traditional information channels warrants further evidence to investigate the interaction between traditional media and social media (e.g., StockTwits) and how various types of information (e.g., *Breaking News, Full Articles, Press Releases)* may influence investors' opinion on StockTwits. To examine the role of *Media Coverage* in influencing *Disagreement* among investors on StockTwits, we sort firms by *Media Coverage* based on each news type. Specifically, for each news type, *Media Coverage* is aggregated monthly and divided into quintiles, where Q1 is the quintile with no/low *Media Coverage*, and Q5 is the quintile with the maximum *Media Coverage* for that specific news type. Using the regression model estimated in Eq. (5), we run a regression for each news type within a specific quintile. Fig. 3 presents the coefficient estimates for our main variable of interest.

In Plot 1, the results from *Overall Media Coverage* show that, when moving from Q1 (no/low media coverage) to Q5 (maximum media coverage), the negative impact of *Disagreement* on *Return Synchronicity* tends to increase. The difference between the coefficients of Q5 and Q1 in Plot 1 is ~10%, which is statistically significant at the 1% level. This suggests that *Media Coverage* allows investors to actively engage in discussion threads on StockTwits after interpreting various types of news. In return, investors on StockTwits increase the diffusion of firm-specific information in financial markets. It is pertinent to note that the impact of *Disagreement* does not vanish even if there is no/low media coverage (the coefficient of Q1 is –3.51%), further illustrating the distinct contribution of social media to predicting firm-specific stock-price variation in addition to traditional information channels. These results are consistent with Roll (1988), who presents evidence in favor of low R^2 in the absence of news. He argues that in addition to the release of public information from news, which is capitalized into stock prices, trading activities of informed arbitrageurs also contribute to the capitalization of private information. Overall, these findings are consistent with the existing literature and highlight the key role of media in allowing investors on StockTwits to consume information and increase the velocity of firm-specific information.

[Insert Figure 3 Here]

We further estimate regressions after segregating media coverage based on three news types. The first type is *Full Articles* written by authors in the finance and investment industry, the second is *Breaking News*, and the third is *Press Releases* issued by the sample firms. Plots 2 and 3 present the *Disagreement* coefficients of *Full Articles* and *Breaking News,* respectively, for each quintile moving from Q1 (no/low media coverage) to Q5 (maximum media coverage) with a 95% confidence interval. The differences between the coefficients of Q5 and Q1 in Plots 2 and 3 are 10.22% and 11.19%, respectively, which is statistically significant at the 1% level. The economic significance of these results implies that *Full Articles and Breaking News* get considerable attention on StockTwits, thus illustrating the impact of *Disagreement* on *Return Synchronicity* and diffusing the flows of firm-specific information in financial markets at a higher rate.

Plot 4 presents the *Disagreement* coefficient for *Press Releases* issued by the firms for each quintile moving from Q1 (no/low media coverage) to Q5 (maximum media coverage) with a 95% confidence interval. These estimates imply that, although statistically significant, *Press Releases* have the least economic significance compared to *Full Articles* and *Breaking News* in moderating the relationship between *Disagreement* and *Return Synchronicity.* The difference between the coefficients for Q5 and Q1 in Plot 3 is 7.35%, at the 1% level of significance, less than those for *Full Articles* and *Breaking News*. Considering the precise and short text nature of *Press Releases*, it may be challenging for investors to interpret information shared via *Press Releases* unless the commentary is already available in the market. Nekrasov et al. (2019) argue that *Press Releases* are a less common tool for engaging with social media audiences by firms unless the same is shared via the firm's social media handles.

We also estimate regressions after segregating media coverage based on *News Topics* and *News Sources*. Our results remain consistent and further highlight how social media and traditional media complement each other to incorporate firm-specific information in financial markets. The regression results are presented in Table A2 of the online Appendix.

4.4. Addressing Endogeneity and Selection Bias

We use two-stage least square (2SLS) regression to address endogeneity concerns and the two-stage Heckman selection model to address self-selection bias.

4.4.1. Two-Stage Least Square Regression

Endogeneity is an important concern in our results for multiple reasons. The recent literature investigating the role of media in financial markets has highlighted the role of selective media coverage (Fang & Peress 2009), sensationalizing rumors (Ahern & Sosyura 2015), and the impact of macro news (Sheng 2019). Similarly, Bhagwat and Burch (2016) present strategic tweeting by firms around their earnings announcements, and Clarke et al. (2020) explore how fake news on social media can influence investors' attention. We use a two-stage least square (2SLS) instrumental variable approach to deal with the endogeneity issue. The instrumental variable approach relies on two main assumptions. First, there should be an independent distribution of the excluded instruments' standard errors, and second, the excluded instruments are highly correlated with the endogenous regressors.

Ivković and Weisbenner (2005) present evidence that investors who invest within 250 miles of their geographic proximity earn 3.2% additional annual returns as compared to their nonlocal investments and suggest that such local bias is information-driven. Similarly, Bodnaruk (2009) highlights the role of the local information effect and presents evidence that diversified investors have better access and expertise to process local information. Consequently, such investors earn higher risk-adjusted returns. Therefore, our first instrument variable is *Proximity,* which captures the social distances between investors and the firms' headquarters for whom they have been discussing and sharing ideas on StockTwits. We use *Proximity* as an instrument variable because it satisfies the criteria of a good instrument; *Proximity* is highly and positively correlated with disagreement as it can be an important source of information due to their close proximity to the firm's headquarters. However, the instrument is unlikely to be correlated with the error term in the second stage regression because it is doubtful that the company's stock performance can directly affect *Proximity*. Our first instrument variable is in line with previous studies by Baloria and Heese (2018), who argue that local newspapers are biased towards firms in close proximity, and Peress (2014), who argues that national newspaper strikes affect the flows of information in financial markets, resulting in a 12% lower trading volume on strike days.

Lee and Mas (2012) argue that there are emerging trends of private unionization and highlight the role of labor unions in the financial markets. They argue that financial markets are slow to react to union actions despite their decremental impact on the firms' equity in the long run. Moreover, Wood and Pasquier (2018) provide evidence that social media play a pivotal role in gaining momentum for labor union activities. Hence, such investor-orientated platforms facilitate workers to share their opinion and gain collective identity. Therefore, we use *Labor_Issues* as an additional instrument for disagreement. *Labor_Issues* is defined as the total number of issues related to firms' labor unions aggregated monthly. Our main intuition is that such *Labor_Issues* is positively associated with disagreement as it can exacerbate the number of discussions on social media platforms. However, it is unlikely that the company's stock performance can directly affect *Labor_Issues.* To ensure our instrument variable's economic significance, we only account for labor union strikes, settlements, and layoffs. Overall, we manually collect 553 *Labor_Issues.* To further verify this instrument variable, we cross-check all these issues using news articles from Factiva. Selected newspaper articles discussing *Labor_Issues* are presented in the Online Appendix.

[Insert Table 5 Here]

Table 5 presents the results from 2SLS regression. The second-stage regression shows the negative association between disagreement and return synchronicity. The results from the first-stage regression present the positive association between disagreement and two instrument variables. These results are in line our previous findings, suggesting that the disagreement among investors provides an inflow of firm-specific information. We also check the validity of our instruments based on the following tests. First, the Sargan test is a test for overidentifying restrictions, which tests for the exclusion condition. A *P-value* of the Sargan test, which is higher than 5% suggest that the excluded instruments are correctly excluded from the estimated equation. For the relevance condition, Stock and Yogo (2005) argue that the weak instruments provide biased instrumental estimators.A rule of thumb for the F-statistic associated with the first stage regression is that it should be greater than 10 (Bound *et al.* 1995; Staiger & Stock 1997). The high value of F-statistics (174.11) suggest the higher explanatory power of our instruments andour two instruments are sufficiently strong to justify inference from the results.

Finally, the additional results from Anderson-Rubin test and the Kleibergen-paap test reject the null hypothesis (*P-values* smaller than 0.05), suggesting that the model is identified. Therefore, the association between endogenous regressors and the instrument variables are adequate to identify the equation.

4.4.2. Two-Stage Heckman Selection Model

Disagreement on StockTwits is a choice for investors, depending on several exogenous factors.[44] However, it is also pertinent to mention that such differences can also be attributed to endogenous factors, since investors on StockTwits can see each other's comments, and popular ideas based on the number of likes and comments can get more attention than the rest. Therefore, such choices can also be determined endogenously. Under such circumstances, self-selection bias could be a potential issue that may influence OLS estimates (Heckman 1979). To address these concerns, Heckman proposes a two-stage model. The first stage is the selection phase, and the second stage is the outcome phase.

We implement the two-stage Heckman (1979) selection model by creating a binary variable to run a probit regression in the selection phase. Our binary variable is the choice between *Disagreement* and *Agreement*. To further ensure the robustness of the model in the first phase, adding an intervening variable that is part of the first stage and not included in the second stage of the model is recommended (Kai & Prabhala 2007). Specifically, this variable should influence our binary variable in the first stage only. We use the *Proximity* additional variable since investors' geographic backgrounds can greatly influence *Disagreement* and Agreement choice. For example, Baik et al. (2016) argue that local Twitter activity can predict higher trading volumes and local social media activity suggest the inflows of private information. In addition to Proximity, we also use the firm-level *R&D/Sales* ratio. This is because firms that allocate more budget to their R&D receive more media coverage. However, it is pertinent to mention that excluding the firm-level *R&D/Sales* ratio does not affect our results in the model's first stage.

Table 5, Model (1) presents the probit regression results for the Heckman selection model's first stage. Our main variable of interest is the association of two additional variables

[44] For example, investors' education, investment type, background, and willingness to participate in different types of communication (e.g., like, post, share, etc.).

with our binary variable. The regression results in Model (1) present that *Proximity* and firm-level *R&D/Sales* ratio can predict *Disagreement,* and it is statistically significant at the 1% level. That is, investors' close proximity to firms' headquarters and firm-level R&D spending can exacerbate the level of *Disagreement* among investors on StockTwits.

From the first stage of the Heckman selection model, we construct an inverse Mills ratio (λ) as an additional regressor to control for self-selection bias in the second phase of the model. Model (4) presents the regression results for the second phase. We find that the Heckman selection model produces qualitatively similar estimates after correcting for a potential selection problem in our sample. Moreover, the coefficients of λ and *Return Synchronicity* are negative and significant at the 1% level, suggesting that certain observed and unobserved factors can increase the likelihood of a higher level of *Disagreement* among investors on StockTwits, further increasing the flows of firm-specific information in financial markets. For example, suppose one interprets the unobserved component as the *Proximity* of investors. In that case, it can be argued that investors with close social proximity may have information that can influence existing investors' opinions and, consequently, they update their recommendations on StockTwits.

4.5. Disagreement and Firm Information Environment

The firm information environment plays a pivotal role in influencing information asymmetry in financial markets and reduces the external financing cost for firms with a transparent information environment (Porta *et al.* 1998; Bushman *et al.* 2004b). A recent study by Bai et al. (2016) argues that the advances in technology since 1960 have increased price informativeness, and financial markets have become more price efficient. However, Nguyen and Kecskés (2020) argue that technology spillovers increase information asymmetry and the complexity of acquiring information in financial markets. Therefore, the existing literature warrants further evidence to understand the firm information environment's role and its implications for social media platforms for investors. We estimate the following regression to understand the moderating effect of the firm information environment:

$$Sync_{i,t} = \beta_0 + \beta_1 Disagreement_{i,t} + \beta_2 Info\_Proxy + \beta_3 Disagreement_{i,t} * Info\_Proxy + \beta_4 Controls + V_i + V_t + V_p + \varepsilon_{i,t} \tag{8}$$

Where *Info_Proxy* is the set of proxies for the information environment. It is pertinent to note that information asymmetry exists among investors and between firms and investors. For example, Hutton et al. (2009a) argue that not all information asymmetry can be associated with exogenous factors. In our context, we are keen to understand the impact of the firm information environment on the level of disagreement among investors on StockTwits. To test this conjecture, we use *Firm Opacity*, *Diversity*, *Industry Concentration,* and *Insider Trading* as different proxies for the firm information environment. The regression results between *Return Synchronicity* and the moderating effect of the firm information environment are presented in Table 6.

[Insert Table 6 Here]

4.5.1. Firm Opacity

Some firms in financial markets are considered opaque as they do not release complete information to financial markets. Lin et al. (2011) argue that firm-information opacity prevents investors from calculating a fair value for the firm and suggest that opaque firms are more likely to face agency issues. Jin and Myers (2006) argue that opaque firms provide less firm-specific information, leaving room for managers to conceal self-serving behaviors. In our case, since disagreement on StockTwits provides firm-specific information, for firms with higher informational opacity, social media platforms for investors, such as StockTwits, can facilitate investors scaling down the further implications of informational opacity by consuming information from StockTwits. We use discretionary accruals (*Disc. Accruals*) as a proxy for *Firm Opacity*. To calculate *Disc. Accruals*, we use Dechow *et al.* (1995) technique and employ a modified Jones (1991) model.[45] We estimate the following cross-sectional regression based on Fama–French 48 industries for each fiscal year and use residuals to calculate $Disc.\ Accruals_{i,t}$ as follows:

$$Disc.\ Accruals_{it} = \frac{TAC_{it}}{Assets_{it-1}} - \left(\widehat{\lambda_0}\frac{1}{Asse\quad_{it-1}} + \widehat{\lambda_1}\frac{\Delta Revenue_{it} - \Delta Receivables_{it}}{Assets_{it-1}} + \widehat{\lambda_2}\frac{PPE_{it}}{Assets_{it-1}}\right) \tag{9}$$

Model (1) presents the interaction between *Disagreement* and *Info_Proxy (*firm opacity). Our variable of interest is the coefficient of the interaction between *Disagreement* and

[45] Total accruals are calculated as income before extraordinary items minus cashflow from operating activities, adjusted for extraordinary items and discontinued operations. Annual data are downloaded from Compustat.

Info_Proxy. This coefficient is statistically significant and negative (–2.40%). These results show that *Firm Opacity* demonstrates the level of *Disagreement* and, consequently, increases the flows of firm-specific information for opaque firms. It is pertinent to note that the stand-alone variable *Info_Proxy* is statistically significant at the 5% level and has a positive coefficient, suggesting that *Firm Opacity* leads to less price-informativeness. These findings suggest that social media platforms for investors such as StockTwits assist investors by offering more insights and analyses to predict stock returns and correctly calculate fair values of opaque firms, which is an otherwise challenging task (Lin *et al.* 2011).

4.5.2. Diversity

Firm-level diversity is defined as the number of business segments and geographic locations in which the firm is operating. Markarian and Parbonetti (2007) argue that diverse firms are complex in nature and face agency issues. This is because diverse firms face challenges in multiple avenues in different geographic locations. Similarly, Bushman et al. (2004a) present evidence that diversity and the governance structure of firms limit the transparency of firm operations to outsiders. They conclude that there is a clear need to improve the corporate transparency of diverse firms since they are complex in nature.[46] Therefore, the flows of firm-specific information for diverse firms can be beneficial for all stakeholders.

In our context, social media platforms for investors can allow stakeholders to access firm-specific information for diverse firms. Following Markarian and Parbonetti (2007), we construct *Diversity* as the natural logarithm of the number of business segments multiplied by the number of geographic segments. Ceteris paribus, the higher the number of business and geographic segments, the higher the uncertainty and demand for information from all stakeholders. Model (2) presents the regression results for *Info_Proxy*. Our main variable of interest is the interaction between *Disagreement* and the *Info_Proxy* variable, i.e., *Diversity*, which is significant at the 1% level with a negative coefficient (–1.8%). This result suggests that disagreement among investors is more pronounced for diverse firms. In contrast, the stand-alone coefficient of Info*_Proxy* (diversity) is insignificant, suggesting that it does not provide any information.

[46] Løwendahl and Revang (1998) highlight the role of the information environment for complex firms and argue that technological changes and sophisticated customers in the post-industrial society have increased firm-level complexity.

4.5.3. Industry Concentration

Ali et al. (2014) argue that firms in highly concentrated industries disclose less information since the cost of information is higher than the utility of the information. This is because, in more concentrated industries, each firm's market share is comparatively higher than in less concentrated industries. Therefore, firms in these industries may provide reliable information to predict the future demands of industry and market trends. However, industry rivals can use that information to prepare a robust future strategy, resulting in intense market competition, thus increasing the proprietary cost of information disclosure for disclosing firms. Verrecchia (1983) shows that firms with a higher proprietary cost of disclosure disclose less information than firms with a lower proprietary cost of information. The less informative disclosure practices in more concentrated industries warrant further evidence to investigate the role of social media platforms for investors such as StockTwits to assist stakeholders by providing firm-specific information.

To test this argument, we created the *Industry Concentration*[47] measure using the Herfindahl-Hirschman index (HHI). We use firms' total assets for the last three years and two-digit standard industrial classification (SIC) codes to calculate HHI. Our results remain consistent when we use firms' total sales instead of total assets. Model (3) presents the regression results for the variable *Info_Proxy*. These results show that the coefficient of the interaction between *Disagreement* and *Info_Proxy* is negative and significant at the 1% level (–1.4%), while the variable *Info_Proxy* is insignificant. Our results provide clear evidence, in line with existing literature, that social media platforms for investors assist stakeholders by providing firm-specific information for firms in more concentrated industries.

4.5.4. Insider Trading

Insider trading activities can affect the firm information environment since such trades are an important source of private information from firms to market participants. Motivated by Piotroski and Roulstone (2004), we investigate the moderating effect of *Insider Trades* on

[47] Ali et al. (2009) provide substantial evidence that Compustat measures of industry concentration provide mixed results with certain limitations. This is because, when measuring industry concentration, one should also consider the overall industry, which includes private firms. They suggest using US Census measures of industry concentration. In our context, this data set is not available with yearly frequency. Our sample period is only five years, while the US Census measure of industry concentration data is only available every five years. However, our findings are consistent with previous literature and using alternative proxies for industry concentration, suggesting that this caveat does not affect our results.

Return Synchronicity. *Insider Trades*[48] are calculated as the absolute value of buy and sell trades scaled by total insider trades in a given month. Model (4) presents the interaction between *Disagreement* and *Info_Proxy*. The coefficient of the interaction is significant at the 1% level and negative (–2.1%), suggesting that higher numbers of insider trades demonstrate the impact of *Disagreement* on *Return Synchronicity* and provide firm-specific information. Given that insiders have an information advantage, disagreement on StockTwits allows stakeholders in financial markets to consume such information and benefit from this advantage, thus reducing information asymmetry under such circumstances.

4.6. Disagreement and Salience

An important aspect of limited attention is the conscious allocation of scarce cognitive resources (Kahneman 1973). Attracting investors' attention in the first place depends on the attention-grabbing characteristics of the stimulus. Such attention-grabbing features are called *Salience* (Fiske & Taylor 2013). In financial markets, *Salience* can be defined as the information itself (*Information signals*[49]), or the *Sources*[50] of information. Such stimuli with different levels of *Salience* compete in financial markets to attract investors' attention. However, only stimuli with higher *Salience* can increase the marginal utility of information acquisition for investors wishing to remain in the competition (Hirshleifer & Teoh 2003; Hirshleifer *et al.* 2011). Previous studies have mainly discussed the impact of limited attention without explaining the effects of *Salience*. However, it plays a vital role in the attention-allocation process. In recent studies by Li et al. (2019) and Huang et al. (2018), they argue that *Salience* is a key feature, based on which investors determine the quality of information signals in financial markets. Therefore, our next strand of investigation is to understand the role of *Salience*. Specifically, our aim is to understand the impact of *Salience* on disagreement among investors on StockTwits. We divide *Salience* into *Information Signals* and the *Heterogeneity of Investors*.

4.6.1. Information Signals

We broadly define the attention-grabbing characteristics of information signals on StockTwits based on their *Network* and *Social Media Attention* (SMA). The *Network* is defined

[48] Data are collected from Thomson Reuters' insider trading database. We use transaction codes P and S, and role codes CB, CEO, CO, GC, and P.
[49] This includes any material information that can be useful for investors to predict future prices and firms' future earnings.
[50] This includes industry professionals and influencers who analyze financial markets.

as the reach of information signals in an extensive social network; i.e., the larger the network, the greater the reach of information signals. Similarly, *SMA* is further divided into three subgroups based on the number of ideas (*Ideas*); the *Popularity* of ideas, which is the number of likes each idea receives; and, finally, *Discussion* on StockTwits, which is defined as the number of replies a specific idea has on StockTwits. We use *Salience* as a proxy to represent all these variables.

[Insert Table 7 Here]

Table 7 presents the regression results for the *Salience* groups, i.e., *Network* and *SMA*, respectively. One of the distinguishing features of social media is its vast networks of users. Because of these extensive social networks, the velocity of information diffusion on social media is far higher than on any other traditional information channel. Model (1) presents the results for the interaction between *Network* and *Disagreement*. Our main variable of interest is the interaction coefficient between *Disagreement* and *Network*, which is negative and significant at the 1% level (–7.8%). The association between *Disagreement* and *Network* further suggests that extensive social networks demonstrate the impact of *Disagreement* on *Return Synchronicity*. The result in Model (1) implies that when influential investors[51] post ideas on StockTwits, this allows others to follow the lead of those ideas across a broader spectrum, consequently increasing the level of *Disagreement* and prompting a higher inflow of firm-specific information in financial markets.

Models (2) to (4) present the regression results for the interaction between *Disagreement* and SMA subgroups, Ideas, Popularity, and Discussion, respectively. All three coefficients of the interactions between these Salience groups and *Disagreement* are negative and statistically significant at the 1% level. It is pertinent to note that attention allocation and responses to any stimuli are simultaneous actions. For example, investors can allocate attention by posting ideas, liking ideas, or participating in discussion threads on StockTwits. However, our results also suggest that posting ideas is the most popular way of allocating attention on StockTwits by comparing the magnitudes of the coefficients associated with the variable of Salience across SMA subgroups, i.e., Ideas, Popularity, and Discussion, based on both non-interacted and interacted with disagreement.

[51] Any investor with a large number of followers on StockTwits is known as an influential investor.

4.6.2. Heterogeneity of Investors

Our final *Salience* group is the *Heterogeneity of Investors* on StockTwits. To examine the impact of the heterogeneity of investors on the level of disagreement among investors, we first define heterogeneity based on the presence of unique investors. Second, we define heterogeneity based on the self-disclosed investors' experience and investment approaches on StockTwits[52]. For example, investors' experience on StockTwits are broadly categorized into three groups: professional, intermediate, and novice. Similarly, investment approaches are categorized into momentum, technical, fundamental, and value[53]. The results are presented in Table 8.

To understand the overall impact of the presence of unique investors, we examine the moderating effect of unique investors on *Disagreement* in Model (1). The coefficient of the interaction term is negative and statistically significant at the 1% level (–7.9%), suggesting that the impact of *Disagreement* on *Return Synchronicity* is more pronounced when *Disagreement* is from more diverse investors. This result is consistent with our previous findings, suggesting that an increase in heterogeneity means higher salience of information signals is attracting more investors to share their ideas on StockTwits.

The next category for the *Heterogeneity of Investors* is investors' self-disclosed experience on StockTwits. For this purpose, we calculate the within-group disagreement among investors who disclose their investment experience on StockTwits. Ideally, professional investors should take the lead to facilitate the flows of firm-specific information as compared to novice investors, and this is what we find in our regression results presented in Model (2). Overall, the professional investors' coefficient magnitude is the largest, followed by intermediate and novice investors. However, we could not reject the null hypothesis of the equality of coefficients of professional and novice investors using the *t*-test, given that the *P*-value is 0.1347. [54]

---

[52] Investors disclose such information voluntarily and these disclosures are not a requirement on StockTwits.
[53] In addition to these investment approaches, investors have the choice to select from global macro and growth investment approaches. However, these investment approaches are less popular. In our sample, there are less than 1% investors who choose such investment approaches. Following Cookson and Niessner (2019), we exclude such approaches. However, including investors with such investment approaches does not change our results.
[54] We follow a conservative approach by using two-dimensional clustering (firm and month). However, when we only use single clustering at the firm level rather than double clustering, the *t*-test of the equality of the coefficients is significant, suggesting that professional investors play an important role in diffusing firm-specific information as compared to novice investors.

Previous studies such as Jegadeesh and Titman (1993) highlight the role of momentum investing and suggest that momentum investors can earn higher returns. Similarly, a recent study by Hillert et al. (2014) suggests that stock covered by media has significantly higher momentum. In our case, StockTwits plays a pivotal role in diffusing the information from various information channels. Therefore, it is important to examine the impact of investment approaches on Return Synchronicity. Model (3) presents regression results using within-group disagreement for momentum, technical, fundamental, and value investors. The results show that within-group disagreement among investors decreases Return Synchronicity, indicating higher inflows of firm-specific information. These results remain consistent across all investment approaches. However, the negative impact of *Disagreement* on Return Synchronicity is higher for Momentum (6.2%), followed by Technical (5.1%) and Fundamental (4.4%). The impact of Value *Disagreement* on Return Synchronicity (2.7%) is less than half of that for Momentum. To further understand the difference between coefficients of *Disagreement* for investment approaches, we conduct the *t*-test for the equality of coefficients between momentum and value investors, and we reject the null hypothesis that two coefficients are equal. It is pertinent to note that we cannot reject the null hypothesis when comparing coefficients of disagreement of Momentum and Technical investors. This is because Momentum investors may also follow Technical investment approaches such as moving averages.

Overall, our results provide compelling evidence that the *Salience* on StockTwits, which comprises *Information Signals* and the *Heterogeneity of Investors,* plays a crucial role in facilitating investors to efficiently allocate their attention and show the flows of firm-specific information in financial markets. These findings are consistent with recent studies suggesting that social interactions play an essential role in influencing investors' behaviors in financial markets (Hirshleifer 2019).

5. Robustness Checks

5.1. Alternative Proxies for Return Synchronicity and Disagreement

In our study, we use Carhart (1997) four-factor model to derive the value of R^2 and then calculate *Return Synchronicity*. As a further check on the robustness of our model, we also use alternative proxies for *Return Synchronicity*. In this vein, Morck *et al.* (2000) model (henceforth MYY) offers a slight variation to derive the value of R^2 as follows:

$$Ret_{i,t} = \lambda_0 + \lambda_1 Market\ Return_{i,t} + \lambda_2 Industry\ Return_{i,t} + \lambda_3 Market\ Return_{i,t-1} + \lambda_4 Industry\ Return_{i,t-1} + \varepsilon_{i,t} \tag{10}$$

Where $Industry\ Return_{i,t}$ is calculated based on two-digit SIC industry codes. In other studies, Peng and Xiong (2006) and Anton and Polk (2014) measure *Return Synchronicity* as a times series of Pearson's correlation coefficients between the firm and market return:

$$CORR_{k,m} = \frac{COV(R_k, R_m)}{\sigma_{R_k}, \sigma_{R_k}} \tag{11}$$

In addition to using alternative proxies, we extend the four-factor and MYY models by using the value of $Adj. R^2$ instead of R^2. This is a conservative approach since a large chunk of sample observations are lost because $Adj. R^2$ cannot be calculated with fewer observations. Our results from the robustness tests using these alternative proxies for *Return Synchronicity* are presented in Table 9. These results are consistent with our model, and our main variable of interest, *Disagreement,* is significant at the 1% level with a negative coefficient, even after using alternative proxies for *Return Synchronicity*.

[Insert Table 9 Here]

5.2. *Alternative Machine-Learning Approaches for Recommendation Predictions and Disagreement*

Machine-learning prediction models are selected based on the quality of their predictions and the availability of parameters that match the data set. In our study, we use the Random Forest Decision Trees (*RFDT*) approach. As a further check on the robustness of our model and to understand if variations in the recommendation predictions model might affect our results, we also predict recommendations using Support Vector Machine (*SVM*) and Maximum Entropy (*MaxEnt*). It is pertinent to note that in all three prediction models, our training data set remains the same, and we use tenfold cross-validation and feature selection. The accuracy and F1 score for SVM are 80% and 87%, respectively. Similarly, the accuracy and F1 score for MaxEnt are 74% and 83%, respectively. The results presented in Table 10 complement our findings, suggesting that when using alternative prediction models, our results remain consistent.

[Insert Table 10 Here]

6. Conclusion

This paper provides new evidence that social media platforms for investors, such as StockTwits, assist investors by providing firm-specific information. More importantly, our study complements the existing literature by offering substantial evidence that social media platforms for investors provide firm-specific information that can help investors make their investment decisions. These findings are consistent with Hong and Stein (2007) heterogenous-agent framework model and existing literature on behavioral finance. Our analysis is based on more than 12 million posts from StockTwits posted by 162,836 unique investors from January 2013 to December 2017, discussing 956 US-listed companies. To predict investors' recommendations and measure the level of disagreement among investors on StockTwits, we use Random Forest Decision Trees as our main prediction model. To the best of our knowledge, this is the first study to highlight the role of discussions on social media in providing firm-specific information.

This study contributes to the existing literature on the role of social media in financial markets in the following areas. First, discussions on StockTwits result in higher inflows of firm-specific information. Second, overall media coverage and investors' recommendation revisions contribute to investors' talk on StockTwits, consequently increasing disagreement and prompting higher inflows of firm-specific information. We extend our investigation to assess the effect of firms' information environment. We find that disagreement among investors on StockTwits plays a pivotal role in the supply of firm-specific information for firms with higher Opacity, Diversity, Industry Concentration, and Insider Trades. Third, using the salience of information signals on StockTwits, we show how investors on social media platforms follow leads from influencers, resulting in new investors opting to participate in discussions. Consequently, such interactions result in higher levels of disagreement and information diffusion in financial markets. These findings are robust to endogeneity, sample selection bias, and using alternative measures of return synchronicity and recommendation predictions.

Our findings have practical implications for portfolio managers and investors. Portfolio managers can develop multiple portfolio strategies based on the level of disagreement and the salience of information signals on social media platforms for investors. After considering the practical implications of investors' opinions on StockTwits, Thomson Reuters Eikon and Bloomberg Terminals have already embedded an online version of StockTwits in their platforms to assist investors and portfolio managers. Unlike other social media platforms for investors, access to StockTwits is free. This motivates investors to sift through any information based on cashtags, investment philosophies, and investment approaches, as well as the period of investment. Our findings complement the emerging literature in social finance by suggesting that the economic significance of social interactions on social media platforms for investors plays a pivotal role in investment decision-making. Finally, this study also contributes to the emerging literature on the role of big data and machine learning in finance by using one of the largest training data sets for recommendation predictions and comparing prediction models to validate the robustness of our results.

References

Acemoglu, D., Chernozhukov, V., Yildiz, M., 2006. Learning and disagreement in an uncertain world. National Bureau of Economic Research (NBER) Working Paper Series, 1-51

Ahern, K.R., Sosyura, D., 2015. Rumor Has It: Sensationalism in Financial Media. Review of Financial Studies 28, 2050-2093

Al-Nasseri, A., Menla Ali, F., 2018. What does investors' online divergence of opinion tell us about stock returns and trading volume? Journal of Business Research 86, 166-178

Ali, A., Klasa, S., Yeung, E., 2009. The Limitations of Industry Concentration Measures Constructed with Compustat Data: Implications for Finance Research. Review of Financial Studies 22, 3839-3871

Ali, A., Klasa, S., Yeung, E., 2014. Industry concentration and corporate disclosure policy. Journal of Accounting and Economics 58, 240-264

An, H., Zhang, T., 2013. Stock price synchronicity, crash risk, and institutional investors. Journal of Corporate Finance 21, 1-15

Anton, M., Polk, C., 2014. Connected stocks. The Journal of Finance 69, 1099-1127

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

Antweiler, W., Frank, M.Z., 2006. Do US stock markets typically overreact to corporate news stories? SSRN Research Paper No. 878091

Armstrong, C.S., Guay, W.R., Weber, J.P., 2010. The role of information and financial reporting in corporate governance and debt contracting. Journal of Accounting and Economics 50, 179-234

Aumann, R.J., 1976. Agreeing to disagree. The Annals of Statistics 4, 1236-1239

Ayers, B.C., Freeman, R.N., 2003. Evidence that analyst following and institutional ownership accelerate the pricing of future earnings. Review of Accounting Studies 8, 47-67

Ayotte, K., 2020. Disagreement and capital structure complexity. The Journal of Legal Studies 49, 1-32

Baik, B., Cao, Q., Choi, S., Kim, J.-M., 2016. Local Twitter activity and stock returns. Available at SSRN 2783670

Baloria, V.P., Heese, J., 2018. The effects of media slant on firm behavior. Journal of Financial Economics 129, 184-202

Banerjee, S., Davis, J., Gondhi, N., 2019. Choosing to disagree in financial markets. SSRN Research Paper No. 3335257

Banerjee, S., Kaniel, R., Kremer, I., 2009. Price drift as an outcome of differences in higher-order beliefs. The Review of Financial Studies 22, 3707-3734

Banerjee, S., Kremer, I., 2010. Disagreement and learning: dynamic patterns of trade. The Journal of Finance 65, 1269-1302

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307-343

Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. Journal of financial economics 75, 283-317

Bhagwat, V., Burch, T.R., 2016. Pump it up? Tweeting to manage investor attention to earnings news. Tweeting to Manage Investor Attention to Earnings News (December 26, 2016)

Bodnaruk, A., 2009. Proximity always matters: Local bias when the set of local companies changes. Review of finance 13, 629-656

Bollen, J., Mao, H.N., Zeng, X.J., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2, 1-8

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American statistical association 90, 443-450

Brockman, P., Liebenberg, I., Schutte, M.G., 2010. Comovement, information production, and the business cycle. Journal of Financial Economics 97, 107-129

Bushman, R., Chen, Q., Engel, E., Smith, A., 2004a. Financial accounting information, organizational complexity and corporate governance systems. Journal of Accounting and Economics 37, 167-201

Bushman, R.M., Piotroski, J.D., Smith, A.J., 2004b. What determines corporate transparency? Journal of accounting research 42, 207-252

Campbell, J.L., DeAngelis, M.D., Moon, J.R., 2019. Skin in the game: personal stock holdings and investors' response to stock analysis on social media. Review of Accounting Studies 24, 731-779

Carhart, M.M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57-82

Carlin, B.I., Longstaff, F.A., Matoba, K., 2014. Disagreement and asset prices. Journal of Financial Economics 114, 226-238

Cen, L., Wei, K.J., Yang, L., 2017. Disagreement, underreaction, and stock returns. Management Science 63, 1214-1231

Chahine, S., Mansi, S., Mazboudi, M., 2015. Media news and earnings management prior to equity offerings. Journal of Corporate Finance 35, 177-195

Chan, K., Hameed, A., 2006. Stock price synchronicity and analyst coverage in emerging markets. Journal of Financial Economics 80, 115-147

Chemmanur, T.J., Yan, A., 2019. Advertising, attention, and stock returns. Quarterly Journal of Finance 9, 1950009

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

Chen, H., Singal, V., Whitelaw, R.F., 2016. Comovement revisited. Journal of Financial Economics 121, 624-644

Chen, J., Hong, H., Stein, J.C., 2002. Breadth of ownership and stock returns. Journal of Financial Economics 66, 171-205

Chue, T.K., Gul, F.A., Mian, G.M., 2019. Aggregate investor sentiment and stock return synchronicity. Journal of Banking & Finance 108, 1-19

Clarke, J., Chen, H., Du, D., Hu, Y., 2020. Fake news, investor attention, and market reaction. Information Systems Research

Cookson, J.A., Niessner, M., 2019. Why don't we agree? evidence from a social network of investors. The Journal of Finance 75, 173-228

Da, Z., Engelberg, J., Gao, P.J., 2011. In Search of Attention. Journal of Finance 66, 1461-1499

Dang, T.L., Dang, M., Hoang, L., Nguyen, L., Phan, H.L., 2020. Media coverage and stock price synchronicity. International Review of Financial Analysis 67

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

Dechow, P.M., Sloan, R.G., Sweeney, A.P., 1995. Detecting earnings management. The Accounting Review 70, 193-225

Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. The Journal of Finance 57, 2113-2141

Dimson, E., 1979. Risk measurement when shares are subject to infrequent trading. Journal of Financial Economics 7, 197-226

Ding, R., Zhou, H., Li, Y., 2019. Social media, financial reporting opacity, and return comovement: evidence from Seeking Alpha. Journal of Financial Markets, 1-25
Drake, M.S., Jennings, J., Roulstone, D.T., Thornock, J.R., 2017. The comovement of investor attention. Management Science 63, 2847-2867
Durnev, A., Morck, R., Yeung, B., 2004. Value‐enhancing capital budgeting and firm‐specific stock return variation. The Journal of Finance 59, 65-105
Durnev, A., Morck, R., Yeung, B., Zarowin, P., 2003. Does greater firm‐specific return variation mean more or less informed stock pricing? Journal of Accounting Research 41, 797-836
Dyck, A., Moss, D., Zingales, L., 2013. Media versus special interests. Journal of Law & Economics 56, 521-553
Dyck, A., Volchkova, N., Zingales, L., 2008. The corporate governance role of the media: evidence from Russia. Journal of Finance 63, 1093-1135
Eyster, E., Rabin, M., Vayanos, D., 2019. Financial markets where traders neglect the informational content of prices. The Journal of Finance 74, 371-399
Fama, E.F., French, K.R., 2007. Disagreement, tastes, and asset prices. Journal of Financial Economics 83, 667-689
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607-636
Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. Journal of Finance 64, 2023-2052
Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? Journal of Machine Learning Research 15, 3133-3181
Firth, M., Lin, C., Wong, S.M., 2008. Leverage and investment under a state-owned bank lending environment: evidence from China. Journal of Corporate Finance 14, 642-653
Fiske, S.T., Taylor, S.E., 2013. Social cognition: From brains to culture. SAGE Publications Ltd.
Giannini, R., Irvine, P., Shu, T., 2019. The convergence and divergence of investors' opinions around earnings news: evidence from a social network. Journal of Financial Markets 42, 94-120
Golez, B., Goyenko, R., 2019. Disagreement in the equity options market and stock returns. SSRN Research Paper No. 3443241
Green, T.C., Hwang, B.-H., 2009. Price-based return comovement. Journal of Financial Economics 93, 37-50
Grullon, G., Kanatas, G., Weston, J.P., 2004. Advertising, breadth of ownership, and liquidity. Review of Financial Studies 17, 439-461
Harris, M., Raviv, A., 1993. Differences of opinion make a horse race. The Review of Financial Studies 6, 473-506
Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica: Journal of the econometric society, 153-161
Hillert, A., Jacobs, H., Müller, S., 2014. Media makes momentum. The Review of Financial Studies 27, 3467-3501
Hirshleifer, D., 2015. Behavioral Finance. Annual Review of Financial Economics 7, 133-159
Hirshleifer, D., 2019. Attention, psychological bias, and social interactions. In: Presentation Slides, for Finance Theory Group Wharton Summer School (Long Version). SSRN
Hirshleifer, D., Lim, S.S., Teoh, S.H., 2011. Limited investor attention and stock market misreactions to accounting information. Review of Asset Pricing Studies 1, 35-73
Hirshleifer, D., Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. Journal of Accounting and Economics 36, 337-386

Hirshleifer, D., Teoh, S.H., 2009. Thought and behavior contagion in capital markets. Handbook of Financial Markets: Dynamics and Evolution, 1-56

Hong, H., Kubik, J.D., Stein, J.C., 2004. Social interaction and stock‐market participation. The journal of finance 59, 137-163

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

Hong, H., Stein, J.C., 2007. Disagreement and the stock market. Journal of Economic Perspectives 21, 109-128

Huang, S., Huang, Y., Lin, T.-C., 2019a. Attention allocation and return co-movement: evidence from repeated natural experiments. Journal of Financial Economics 132, 369-383

Huang, S., Maharjan, J., Thakor, A.V., 2019b. Disagreement-induced CEO turnover. Journal of Financial Intermediation, 100819

Huang, X., Nekrasov, A., Teoh, S.H., 2018. Headline salience, managerial opportunism, and over-and underreactions to earnings. The Accounting Review 93, 231-255

Hutton, A.P., Marcus, A.J., Tehranian, H., 2009a. Opaque financial reports, R2, and crash risk. Journal of financial Economics 94, 67-86

Hutton, A.P., Marcus, A.J., Tehranian, H., 2009b. Opaque financial reports, R-2, and crash risk. Journal of Financial Economics 94, 67-86

Ivković, Z., Weisbenner, S., 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. The Journal of Finance 60, 267-306

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of finance 48, 65-91

Jiang, L., Liu, J., Yang, B., 2019. Communication and comovement: evidence from online stock forums. Financial Management 48, 805-847

Jin, L., Myers, S.C., 2006. R2 around the world: new theory and new tests. Journal of Financial Economics 79, 257-292

Jones, J.J., 1991. Earnings management during import relief investigations. Journal of Accounting Research 29, 193-228

Kahneman, D., 1973. Attention and effort. Prentice-Hall, Inc., Englewood Cliffs,New Jersey.

Kai, L., Prabhala, N.R., 2007. Self-selection models in corporate finance. In: Handbook of empirical corporate finance. Elsevier, pp. 37-86.

Kandel, E., Pearson, N.D., 1995. Differential interpretation of public signals and trade in speculative markets. Journal of Political Economy 103, 831-872

Karpoff, J.M., 1986. A theory of trading volume. The journal of finance 41, 1069-1087

Lee, D.S., Mas, A., 2012. Long-run impacts of unions on firms: New evidence from financial markets, 1961–1999. The Quarterly Journal of Economics 127, 333-378

Li, Q., Lourie, B., Teoh, S.H., 2019. How salience of management guidance affects forecasting behavior: evidence from a quasi-natural experiment on estimize. SSRN Research Paper No. 3423608

Lin, C., Ma, Y., Xuan, Y., 2011. Ownership structure and financial constraints: evidence from a structural estimation. Journal of Financial Economics 102, 416-431

Løwendahl, B., Revang, Ø., 1998. Challenges to existing strategy theory in a postindustrial society. Strategic management journal 19, 755-773

Markarian, G., Parbonetti, A., 2007. Firm complexity and board of director composition. Corporate governance: an international review 15, 1224-1243

Mathers, A.M., Wang, B., Wang, X.S., 2017. Innovation and price informativeness. Financial Management 46, 523-546

Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. The Journal of Finance 32, 1151-1168

Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? Journal of Financial Economics 58, 215-260

Nguyen, P.-A., Kecskés, A., 2020. Do technology spillovers affect the corporate information environment? Journal of Corporate Finance, 101581

Odean, T., 1998. Volume, volatility, price, and profit when all traders are above average. The journal of finance 53, 1887-1934

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80, 563-602

Peress, J., 2014. The media and the diffusion of information in financial markets: evidence from newspaper strikes. The Journal of Finance 69, 2007-2043

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. The Review of Financial Studies 22, 435-480

Pindyck, R.S., Rotemberg, J.J., 1993. The comovement of stock prices. The Quarterly Journal of Economics 108, 1073-1104

Piotroski, J.D., Roulstone, D.T., 2004. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. The Accounting Review 79, 1119-1151

Porta, R.L., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.W., 1998. Law and finance. Journal of political economy 106, 1113-1155

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

Roll, R., 1988. R2. The Journal of Finance 43, 541-566

Sheng, J., 2019. Macro news, micro news, and stock prices. In: 44th Annual Meeting of the European Finance Association. University of Mannheim, Mannheim, Germany

Shiller, R.J., 1989. Comovements in stock prices and comovements in dividends. The Journal of Finance 44, 719-729

Shiller, R.J., 1992. Market volatility. MIT press.

Sila, V., Gonzalez, A., Hagendorff, J., 2017. Independent director reputation incentives and stock price informativeness. Journal of Corporate Finance 47, 219-235

Sprenger, T.O., Tumasjan, A., Sandner, P.G., Welpe, I.M., 2014. Tweets and trades: the information content of stock microblogs. European Financial Management 20, 926-957

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica: journal of the Econometric Society, 557-586

Stock, J., Yogo, M., 2005. Asymptotic distributions of instrumental variables statistics with many instruments. Identification and inference for econometric models: Essays in honor of Thomas Rothenberg, 109-120

Varian, H.R., 1985. Divergence of opinion in complete markets: a note. The Journal of Finance 40, 309-317

Varian, H.R., 1989. Differences of opinion in financial markets. In: financial Risk: Theory, Evidence and implications. Springer, pp. 3-37.

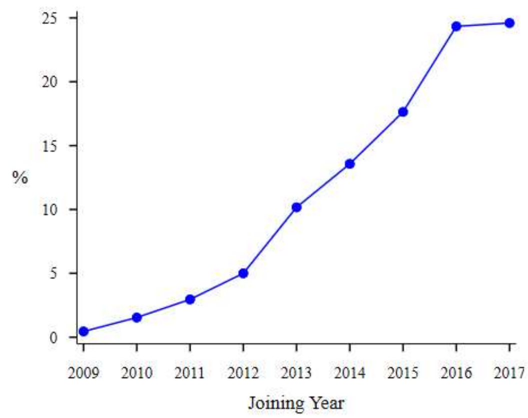
Vega, C., 2006. Stock price reaction to public and private information. Journal of Financial Economics 82, 103-133

Veldkamp, L.L., 2006. Information markets and the comovement of asset prices. The Review of Economic Studies 73, 823-845

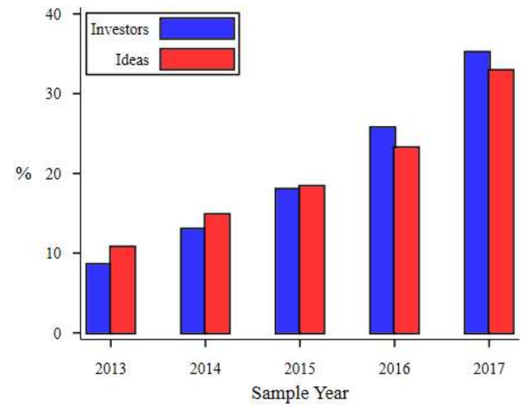
Verrecchia, R.E., 1983. Discretionary disclosure. Journal of accounting and economics 5, 179-194

Vives, X., Yang, L., 2017. Costly interpretation of asset prices. CEPR Discussion Paper No. DP12360
Wood, A., Pasquier, V., 2018. The power of social media as a labour campaigning tool: lessons from Our Walmart and the Fight for 15. Pasquier V and Wood AJ (2018) ETUI Policy Brief, European Economic, Employment and Social Policy
Yu, J., 2011. Disagreement and return predictability of stock portfolios. Journal of Financial Economics 99, 162-183

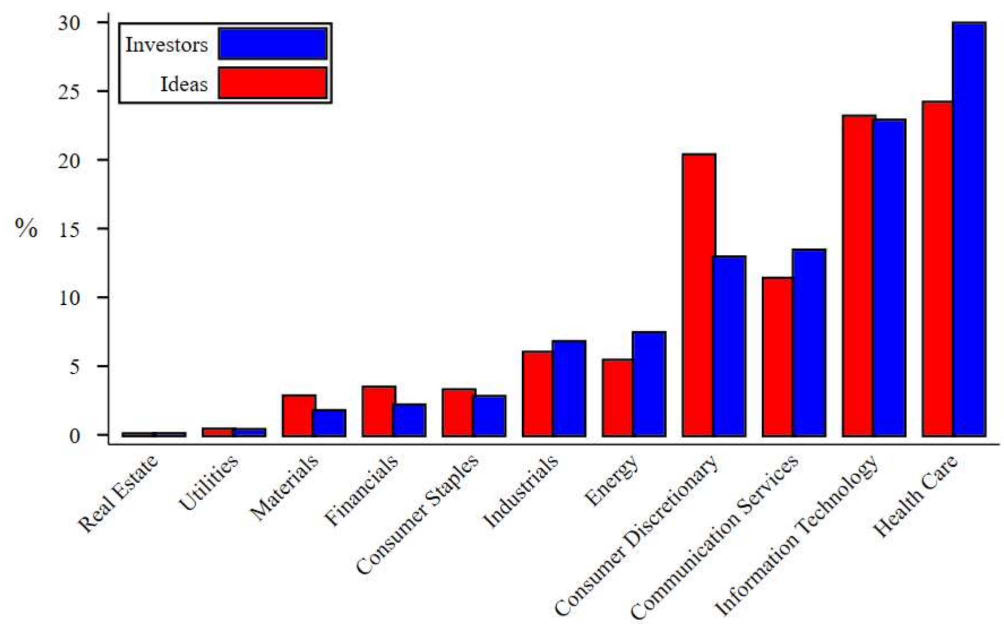
Figure 1: Distribution of Investors and Ideas on StockTwits



Panel A: Distribution of Investors



Panel B: Distribution of Ideas and Investors



Panel C: Distribution of Ideas and Investors across GICS Sectors

Notes: Fig. 1 presents the distribution of ideas and investors in the StockTwits sample. Panel A presents the distribution of investors based on the year they joined StockTwits. Panel B presents the distribution of investors and ideas posted by these investors on StockTwits during the sample period. Panel C presents the distribution of investors and ideas based on the Global Industrial Classification System (GICS) sectors.

Figure 2a: Distribution of StockTwits' ideas across the USA

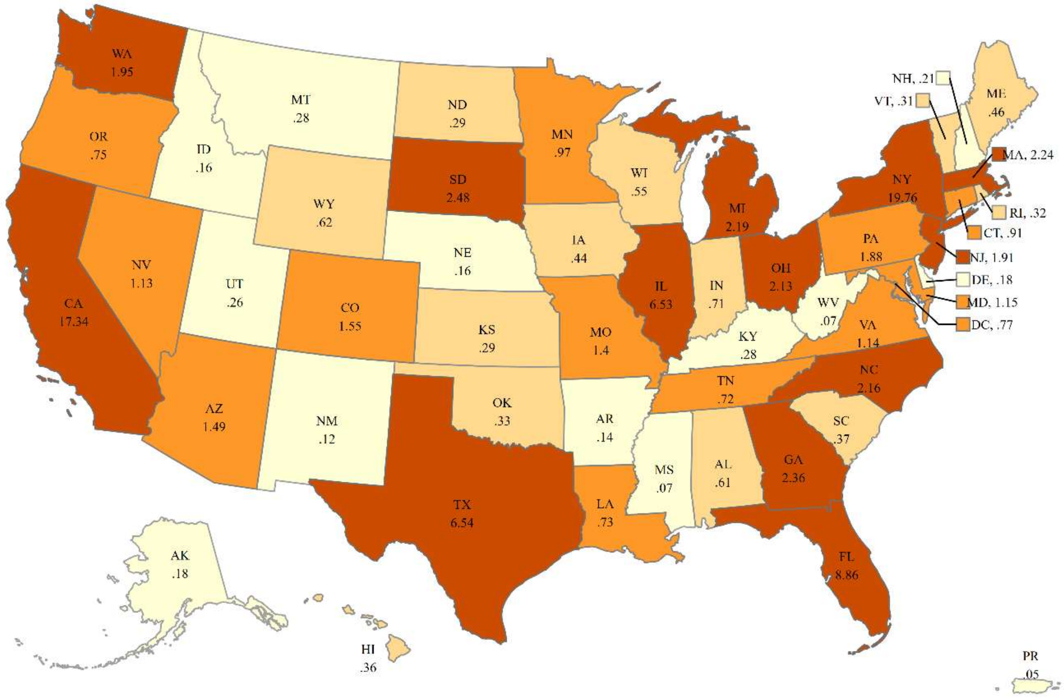
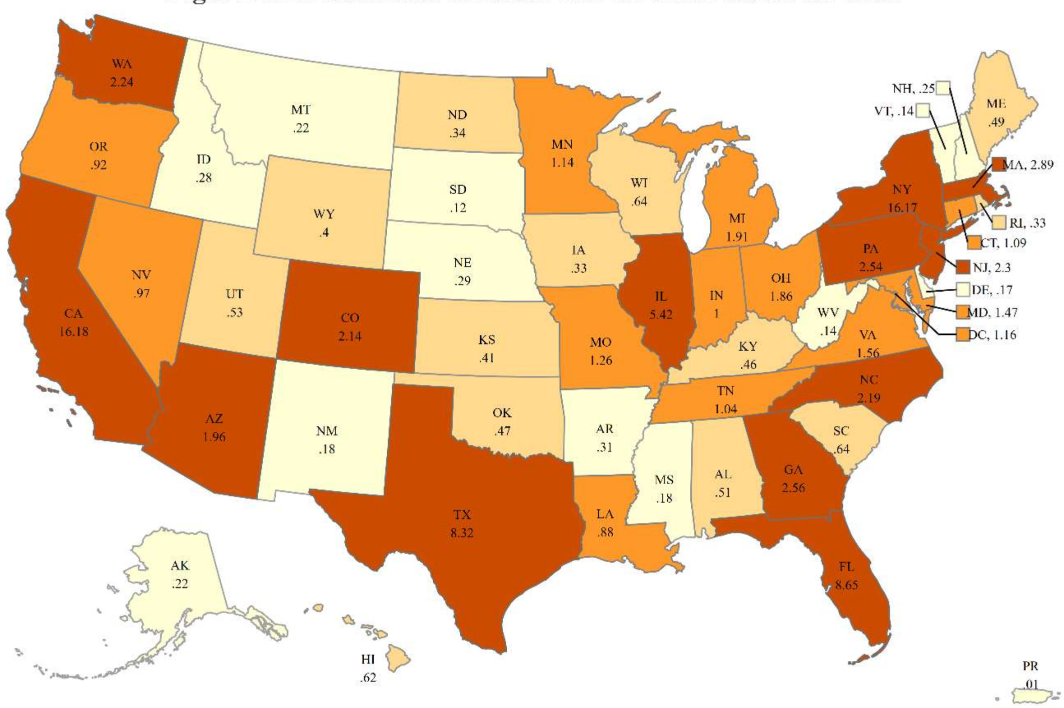
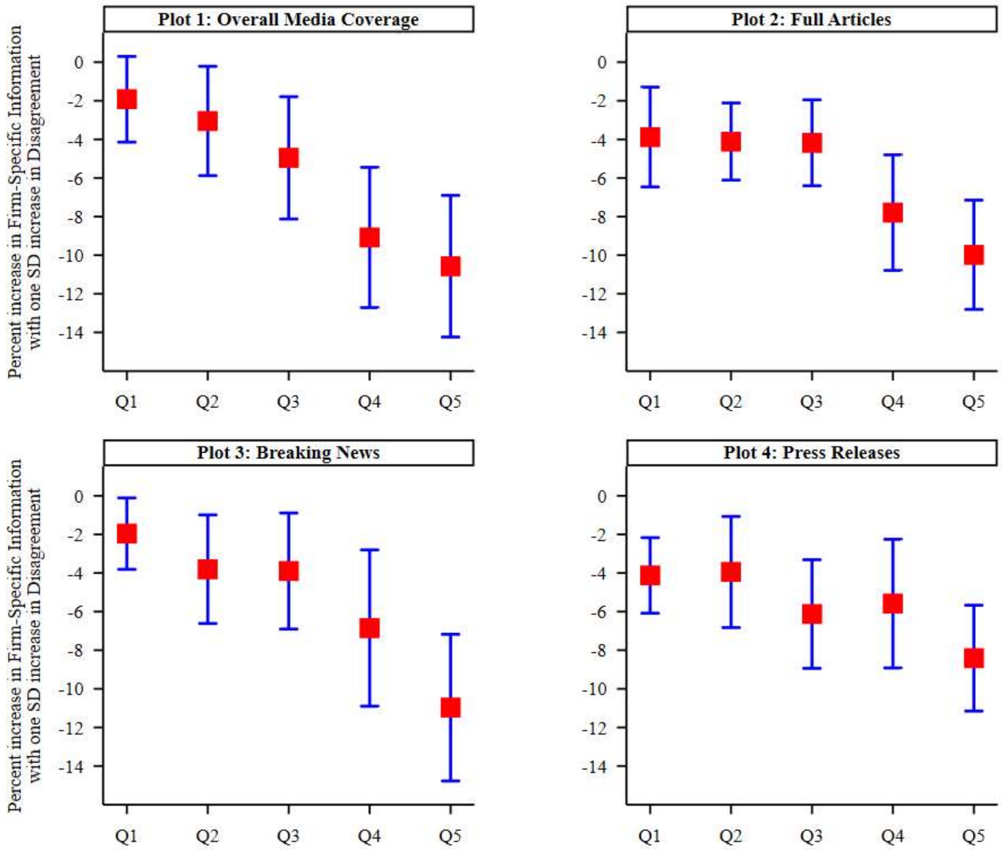


Figure 2b: Distribution of StockTwits' investors across the USA



Notes: These figures present the distribution of StockTwits' ideas and investors across the USA, respectively. Users' locations are collected from their StockTwits public profiles. Users' locations are further cleaned using text analysis and then matched with US state-level coordinates obtained from US Census website geographic data.

Figure 3: Disagreement & Media Coverage



Notes: The figure presents coefficient estimates from the multivariate fixed-effect regression between Disagreement and Return Synchronicity along with their corresponding 95% confidence intervals. The dependent variable is Return Synchronicity, which is calculated using Carhart's (1997) four-factor model. Disagreement is derived from users' recommendations on StockTwits. Media coverage is segregated into groups based on the type of information. Each group is divided into quintiles based on the number of news articles published monthly, where quintile 1 presents firm-month observations with no/low media coverage, and quintile 5 presents firm-month observations with the highest media coverage. For each quintile, the standard regression model in equation (5) is used after excluding media coverage as an explanatory variable. The type groups are defined as the number of press releases issued by the sample firm, full articles as the number of detailed articles that discuss the sample firms, and breaking news as the number of news flashes that explicitly mention the sample firm. The regressions are estimated using time, firm, and industry fixed-effects. The sample consists of 956 firms with 52,888 firm-month observations for the sample period of 2013–2017. Variable definitions are presented in Appendix A.

Table 1: Summary Statistics

Notes: The table reports the summary statistics of StockTwits' ideas and investor-level information, media coverage, and firm-level characteristics of the sample firms. All the variables are defined in Appendix A. Disagreement, and Revisions are calculated using investors' recommendations on StockTwits. These recommendations are predicted using Random Forest Decision Trees (RFDT) method. Further details about this method are discussed in Appendix A.

	Mean	SD	P10	P25	Median	P75	MAX
Panel A: StockTwits Ideas and Investor Level Information							
Disagreement	0.53	0.26	0	0.43	0.59	0.7	1
Revisions	3.61	2.87	0	1.39	3.22	5.21	17.27
Network	12.97	2.18	11.08	12.35	13.28	14.13	18.27
Investors	3.75	1.30	2.20	2.94	3.69	4.49	9.55
Ideas	3.99	1.44	2.30	3.09	3.89	4.78	10.81
Popularity	2.06	1.98	0	0	1.61	3.22	10.90
Discussion	2.60	2.34	0	0	2.30	4.26	12.36
Proximity	0.64	1	0	0	0	1.10	7.79
Panel B: Media Coverage							
	Mean	**SD**	**P10**	**P25**	**Median**	**P75**	**MAX**
Media Coverage (Overall)	2.95	1.06	1.79	2.3	2.94	3.56	8.55
Breaking News	1.61	1.14	0	0.69	1.61	2.4	6.07
Full Articles	1.85	1.22	0	1.10	1.79	2.64	7.44
Press Release	1.61	1.17	0	0.69	1.61	2.4	6.31
Panel C: Firm Level Characteristics							
	Mean	**SD**	**P10**	**P25**	**Median**	**P75**	**MAX**
Return Synchronicity	-0.37	1.04	-1.70	-1.02	-0.33	0.33	3.85
Firm Size	19.33	51.57	0.06	0.59	3.21	15.48	867.51
Analyst Coverage	12.07	8.29	2	5	11	17	54
Leverage	0.25	0.25	0	0.04	0.21	0.37	3.44
Market/Book Ratio	4.32	6.37	0.41	1.36	2.63	4.92	110.53
Adv/Sales	0.02	0.22	0	0	0	0.02	11.36
Sales Growth	0.60	0.42	0	0.49	0.69	0.75	6.83
ROA	0.04	0.30	-0.33	0.02	0.10	0.17	1.42
Earnings Volatility	0.07	0.14	0.01	0.01	0.03	0.08	5.19
Real GDP	0.02	0.01	0.01	0.02	0.02	0.03	0.04
Firm Opacity	-0.24	0.28	-0.59	-0.39	-0.18	-0.06	2.16
Diversity	1.66	1.15	0	0.69	1.95	2.71	4.29
Competition	0.01	0.02	0	0	0	0.01	0.35
Insider Trading	0.45	0.49	0	0	0	1	1

Table 2: Return Synchronicity and Disagreement

Notes: The table reports the regression results of Disagreement and Return Synchronicity. The dependent variable is Return Synchronicity, which is calculated using the Carhart (1997) four-factor model. Disagreement is derived from the investors' recommendations on StockTwits. All right-hand side variables are standardised. The sample consists of 956 firms with 53,778 firm-month observations for the sample period between 2013 - 2017. The regressions are estimated using ordinary least square (OLS) method with time, firm, and industry fixed effects in Model (1) – (3). However, in Model (4) and (5) regressions are estimated using OLS and Fama McBeth regression without fixed effects, respectively. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm and month levels in Model (1) – (4). In Model (5), standard errors are adjusted for cross-sectional dependence based on Fama-McBeth regression.

	OLS				**Fama-MacBeth**
	1	**2**	**3**	**4**	**5**
Disagreement	-0.069***	-0.054***	-0.057***	-0.079***	-0.075***
	[0.0119]	[0.0117]	[0.0109]	[0.0120]	[0.0064]
Media Coverage		-0.221***	-0.227***	-0.000	-0.050***
		[0.0266]	[0.0237]	[0.0190]	[0.0104]
Analyst Coverage			0.116***	0.256***	0.199***
			[0.0145]	[0.0168]	[0.0144]
Leverage			-0.095***	-0.027**	-0.022***
			[0.0238]	[0.0108]	[0.0064]
Adv/Sales			-0.015*	-0.030***	-0.081***
			[0.0071]	[0.0096]	[0.0146]
Market/Book Ratio			-0.032**	-0.026**	-0.027***
			[0.0112]	[0.0109]	[0.0085]
Firm Size			0.203***	0.096***	0.281***
			[0.0578]	[0.0249]	[0.0580]
ROA			0.053***	0.017	-0.009
			[0.0160]	[0.0150]	[0.0097]
Earnings Volatility			0.015**	-0.033*	-0.052***
			[0.0055]	[0.0162]	[0.0083]
Sales Growth			0.052	0.099***	0.086***
			[0.0302]	[0.0280]	[0.0148]
Real GDP $_{t-1}$			-0.075*	-0.072	-0.001
			[0.0410]	[0.0412]	[0.0506]
Fixed Effects	Y	Y	Y	N	N
Adj/Avg. R-squared	0.195	0.211	0.228	0.089	0.128/0.138
Firms	956	956	956	956	956
Observations	53,778	53,778	52,888	52,888	52,888

Table 3: Disagreement and Price Informativeness

Note: The table reports the regression results of the Disagreement and the moderating effect of price informativeness. The dependent variable is cumulative abnormal returns (CAR) calculated quarterly. To calculate CAR, the benchmark return is calculated based on 5 x 5 Size and Book to Market portfolios. Change in earnings ($\Delta Earnings_t$) is used as a proxy of price informativeness and is calculated as firms' earnings at time t minus firms' earnings at time t-1 divided by the beginning of the time t market value equity of the firm. Disagreement is derived from the investors' recommendations on StockTwits and calculated at a quarterly frequency. The sample consists of 949 firms with 17,309 firm-quarter observations for the sample period between 2013 - 2017. The regressions are estimated using ordinary least square (OLS) method with firm fixed effects in Model (1) – (3). However, in Model (4) - (6) and (7) - (9) regressions are estimated using OLS and Fama-McBeth regression without fixed effects, respectively. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm levels in Model (1) – (6). In Model (7) – (9), standard errors are adjusted for cross-sectional dependence based on Fama-McBeth regression.

	OLS						Fama-MacBeth		
	1	2	3	4	5	6	7	8	9
Disagreement		-0.196***	-0.185***		-0.164***	-0.190***		0.360	0.170
		[0.0119]	[0.0114]		[0.0107]	[0.0104]		[0.5211]	[0.4483]
$\Delta Earnings_{t+1}$	-0.001	-0.016***	-0.015***	-0.002	-0.012**	-0.012**	0.120**	-0.111	-0.118
	[0.0015]	[0.0054]	[0.0058]	[0.0014]	[0.0058]	[0.0057]	[0.0499]	[0.1123]	[0.1091]
*Disagreement * $\Delta Earnings_{t+1}$*		0.021***	0.022***		0.015*	0.015**		0.385*	0.390**
		[0.0070]	[0.0070]		[0.0077]	[0.0074]		[0.1924]	[0.1839]
$\Delta Earnings_t$	0.006***	0.008	0.005	0.006***	0.009*	0.008	0.030	-0.001	-0.049
	[0.0019]	[0.0048]	[0.0047]	[0.0021]	[0.0049]	[0.0049]	[0.0378]	[0.1761]	[0.1650]
*Disagreement * $\Delta Earnings_t$*		-0.004	-0.002		-0.005	-0.003		0.030	0.098
		[0.0081]	[0.0079]		[0.0090]	[0.0087]		[0.3055]	[0.2852]
$\Delta Earnings_{t-1}$	-0.002	0.011***	0.007*	-0.002	0.008*	0.007	0.058**	0.150	0.120
	[0.0018]	[0.0040]	[0.0038]	[0.0017]	[0.0044]	[0.0043]	[0.0214]	[0.1067]	[0.1001]
*Disagreement * $\Delta Earnings_{t-1}$*		-0.023***	-0.018***		-0.017***	-0.015**		-0.153	-0.116
		[0.0056]	[0.0053]		[0.0065]	[0.0061]		[0.1749]	[0.1619]
Controls	N	N	Y	N	N	Y	N	N	Y
Fixed Effects	Y	Y	Y	N	N	N	N	N	N
Adj/Avg. R-squared	0.036	0.057	0.095	0.004	0.022	0.049	0.012/0.014	0.044/0.051	0.083/0.091
Firms	946	946	946	949	949	949	949	949	949
Observations	17,306	17,306	17,306	17,309	17,309	17,309	17,309	17,309	17,309

Table 4: Return Synchronicity, Disagreement and Revisions

Notes: The table reports the regression results of Disagreement and Recommendation Revisions of investors on StockTwits. The dependent variable is Return Synchronicity, which is calculated using the Carhart (1997) four-factor model. Disagreement is derived from the investors' recommendations on StockTwits. Recommendation revisions are calculated as the number of revisions from Bullish-Bearish and vice versa by each investor at time *t* to *t-1* and aggregated at a monthly frequency. All right-hand side variables are standardised. The sample consists of 956 firms with 53,778 firm-month observations for the sample period between 2013 - 2017. The regressions are estimated using time, firm, and industry fixed effects. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm and month levels.

	1	2	3
Disagreement		0.010	-0.051***
		[0.0114]	[0.0094]
Revisions	-0.217***	-0.230***	
	[0.0212]	[0.0229]	
*Disagreement * Revisions* $_t$		-0.025*	
		[0.0118]	
Revisions $_{t-1}$	-0.020**		-0.025**
	[0.0084]		[0.0084]
*Disagreement * Revisions* $_{t-1}$			-0.003
			[0.0036]
Media Coverage	-0.138***	-0.141***	-0.218***
	[0.0226]	[0.0227]	[0.0239]
Analyst Coverage	0.113***	0.112***	0.115***
	[0.0151]	[0.0146]	[0.0146]
Leverage	-0.037	-0.038	-0.092***
	[0.0232]	[0.0232]	[0.0233]
Adv/Sales	-0.014**	-0.013*	-0.016**
	[0.0064]	[0.0070]	[0.0064]
Market/Book Ratio	-0.044***	-0.039***	-0.038***
	[0.0112]	[0.0109]	[0.0111]
Firm Size	0.168**	0.173**	0.193***
	[0.0560]	[0.0559]	[0.0574]
ROA	0.057***	0.054***	0.056***
	[0.0148]	[0.0147]	[0.0159]
Earnings Volatility	0.015**	0.015**	0.016**
	[0.0051]	[0.0051]	[0.0060]
Sales Growth	0.076**	0.070*	0.057*
	[0.0322]	[0.0321]	[0.0297]
Real GDP $_{t-1}$	-0.073*	-0.068	-0.082*
	[0.0392]	[0.0398]	[0.0401]
Fixed Effects	Y	Y	Y
Adj. R-squared	0.249	0.248	0.230
Firms	956	956	956
Observations	52,888	52,888	52,888

Table 5: Testing for Endogeneity and Selection Bias

Notes: The table reports the results from two-stage least square (2SLS) regression using the instrumental variable approach and two-stage Heckman Selection model. There are two instruments in the 2SLS regression. Proximity is defined as the number of investors who post ideas on StockTwits while discussing the sample firms and who have the same US state where the firms' headquarter is located. Labor_Issues, defined as the total number of issues related to firms' labor unions aggregated monthly. Sanderson-Windmeijer F test of excluded instruments is presented as S-W F-statistics, Kleibergen-Paap Wald F statistic is presented as K-P Wald F-statistics, and Anderson-Rubin Wald test is presented as A-R Wald F-statistics. In the two-stage Heckman selection model, the first stage selection equation is estimated by probit regression, where the dependent variable is 1 in case of Disagreement and 0 otherwise. In the second stage, the dependent variable is the Return Synchronicity, and the inverse mills ratio (λ) adjusts for the nonzero mean of error terms. All right-hand side variables are standardized. The sample consists of 956 firms with 52,888 firm-month observations for the sample period between 2013 - 2017. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm level.

	2SLS Regression		**Heckman Selection**	
	1	**2**	**3**	**4**
	First Stage	**Second Stage**	**Selection**	**Outcome**
Disagreement		-0.721***		-0.029***
		[0.066]		[0.008]
Proximity	0.168***		0.579***	
	[0.009]		[0.013]	
Labour_Issues	0.009***			
	[0.002]			
R&D/Sales			0.251***	
			[0.097]	
Media Coverage	0.126***	-0.120***		-0.031***
	[0.008]	[0.014]		[0.006]
Firm Size	0.011	0.204***	-0.011	0.566***
	[0.014]	[0.020]	[0.007]	[0.017]
Analyst Coverage	0.070***	0.162***	0.214***	
	[0.014]	[0.016]	[0.007]	
Leverage	0.131***	0.004		-0.009*
	[0.020]	[0.025]		[0.005]
Market/Book Ratio	-0.066***	-0.077***		0.035***
	[0.013]	[0.014]		[0.005]
Adv/Sales	0.000	-0.014**	-0.002	-0.019***
	[0.004]	[0.006]	[0.009]	[0.004]
Sales Growth	0.105***	0.126***	0.216***	0.066***
	[0.027]	[0.036]	[0.018]	[0.012]
ROA	0.005	0.058***	0.017**	0.046***
	[0.013]	[0.017]	[0.009]	[0.006]
Earnings Volatility	-0.006	0.011	0.049***	-0.029***
	[0.008]	[0.008]	[0.009]	[0.006]
Real GDP $_{t-1}$	0.008	-0.066***		-0.071***
	[0.005]	[0.006]		[0.005]
λ				-0.098***
				[0.033]
Fixed Effects	Y	Y	N	N
Firms	956	956	956	956
Observations	52,888	52,888	53,221	52,888
S-W F-statistics	174.11***			
K–P Wald F-statistic	174.11***			
A-R Wald F-statistics	86.68***			
Sargan *P-Value*		0.345		

Table 6: Disagreement and Firm information environment

Notes: The table reports the regression results of the Disagreement and the moderating effect of the firm's information environment. Variable Info_Proxy is the proxy of variables from the firm information environment. The dependent variable is Return Synchronicity, which is calculated using the Carhart (1997) four-factor model. Disagreement is derived from the investors' recommendations on StockTwits. Firm opacity is calculated as a measure of firms' accrual quality based on extended Jones (1991) model. Diversity is the natural logarithm of the number of business segments of the firm multiplied by the number of geographic segments. Competition is based on firm level assets calculated as a proxy of industry competition using the Herfindahl-Hirschman Index. Insider trading is calculated as the absolute value of the difference between buying and selling insider trades scaled by the total insider trades in a given month of a sample firm. All right-hand side variables are standardised. The sample consists of 956 firms with 49,723 firm-month observations in the Model (1) and 52,888 firm-month observations in the Model (2) – (4) for the sample period between 2013 - 2017. The regressions are estimated using time, firm, and industry fixed effects. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm and month levels.

	1	**2**	**3**	**4**
	Firm Opacity	**Diversity**	**Ind. Concentration**	**Insider Trading**
Disagreement	-0.052***	-0.047***	-0.046***	-0.058***
	[0.0074]	[0.0071]	[0.0075]	[0.0106]
Info_Proxy	0.035**	-0.013	-0.160	0.017*
	[0.0128]	[0.0193]	[0.0904]	[0.0087]
*Disagreement * Info_Proxy*	-0.024***	-0.018***	-0.014***	-0.021***
	[0.0050]	[0.0055]	[0.0044]	[0.0053]
Media Coverage	-0.232***	-0.233***	-0.234***	-0.234***
	[0.0201]	[0.0199]	[0.0204]	[0.0247]
Analyst Coverage	0.100***	0.099***	0.106***	0.115***
	[0.0125]	[0.0096]	[0.0122]	[0.0144]
Leverage	-0.056**	-0.061**	-0.055**	-0.094***
	[0.0226]	[0.0214]	[0.0218]	[0.0238]
Adv/Sales	-0.015**	-0.016**	-0.015**	-0.015*
	[0.0064]	[0.0060]	[0.0060]	[0.0070]
Market/Book Ratio	-0.006	-0.010	-0.004	-0.033**
	[0.0077]	[0.0072]	[0.0075]	[0.0110]
Firm Size	0.046*	0.048*	0.049*	0.205***
	[0.0240]	[0.0238]	[0.0243]	[0.0576]
ROA	0.044**	0.050***	0.053***	0.053***
	[0.0152]	[0.0149]	[0.0153]	[0.0160]
Earnings Volatility	0.021**	0.022**	0.022**	0.015**
	[0.0091]	[0.0088]	[0.0088]	[0.0055]
Sales Growth	0.031	0.044*	0.044*	0.052
	[0.0208]	[0.0211]	[0.0209]	[0.0301]
Real GDP $_{t-1}$	-0.144**	-0.146**	-0.147**	-0.076*
	[0.0595]	[0.0590]	[0.0596]	[0.0410]
Fixed Effects	Y	Y	Y	Y
Adj. R-squared	0.226	0.254	0.257	0.228
Firms	918	956	956	956
Observations	49,723	52,888	52,888	52,888

Table 7: Disagreement and Salience of information signals

Notes: The table reports the regression results of the Disagreement and the moderating effect of the salience of information signals of StockTwits. Salience is divided into two groups, i.e., Reach, which represents the magnitude of access to ideas posted by investors on StockTwits; Social Media Attention (SMA) represents alternative attention proxies of StockTwits. These salience groups are further divided into different subgroups, which are defined in appendix A. The dependent variable is Return Synchronicity, which is calculated using the Carhart (1997) four-factor model. Disagreement is derived from the investors' recommendations on StockTwits. All right-hand side variables are standardised. The sample consists of 956 firms with overall 52,888 firm-month observations for the sample period between 2013 - 2017. The regressions are estimated using time, firm, and industry fixed effects. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm and month levels.

	1	SMA		
		2	**3**	**4**
	Network	**Ideas**	**Popularity**	**Discussion**
Disagreement	-0.067***	-0.059***	-0.066***	-0.046***
	[0.0120]	[0.0120]	[0.0102]	[0.0099]
Salience	-0.157***	-0.208***	-0.120***	-0.132***
	[0.0174]	[0.0235]	[0.0228]	[0.0171]
*Disagreement * Salience*	-0.078***	-0.075***	-0.054***	-0.026***
	[0.0084]	[0.0093]	[0.0146]	[0.0084]
Media Coverage	-0.174***	-0.148***	-0.199***	-0.196***
	[0.0225]	[0.0226]	[0.0227]	[0.0237]
Analyst Coverage	0.120***	0.112***	0.115***	0.115***
	[0.0150]	[0.0148]	[0.0147]	[0.0144]
Leverage	-0.087***	-0.034	-0.050**	-0.066**
	[0.0236]	[0.0233]	[0.0213]	[0.0242]
Adv/Sales	-0.013*	-0.013*	-0.015*	-0.015*
	[0.0071]	[0.0072]	[0.0075]	[0.0072]
Market/Book Ratio	-0.031**	-0.040***	-0.041***	-0.036***
	[0.0112]	[0.0108]	[0.0109]	[0.0112]
Firm Size	0.195***	0.182***	0.197***	0.195***
	[0.0590]	[0.0563]	[0.0574]	[0.0561]
ROA	0.053***	0.054***	0.056***	0.054***
	[0.0156]	[0.0149]	[0.0153]	[0.0154]
Earnings Volatility	0.016**	0.014**	0.014**	0.015**
	[0.0052]	[0.0049]	[0.0052]	[0.0053]
Sales Growth	0.056*	0.074**	0.069*	0.059*
	[0.0303]	[0.0325]	[0.0314]	[0.0317]
Real GDP $_{t-1}$	-0.071	-0.069	-0.078*	-0.073*
	[0.0407]	[0.0397]	[0.0406]	[0.0402]
Fixed Effects	Y	Y	Y	Y
Adj. R-squared	0.235	0.247	0.237	0.236
Firms	956	956	956	956
Observations	52,888	52,888	52,888	52,888

Table 8: Disagreement and Heterogeneity of Investors

Notes: The table reports the regression results of the Disagreement and the moderating effect of the heterogeneity of investors on StockTwits. The dependent variable is Return Synchronicity, which is calculated using the Carhart (1997) four-factor model. Disagreement is derived from the investors' recommendations on StockTwits. The heterogeneity of investors is further divided based on the investors' experience and investment approaches. Model (1) presents the regression results based on overall investors, where investors are defined as the number unique of investors posting ideas on StockTwits while discussing the sample firms. Model (2) and (3) presents the regression results based on within group disagreement between investors with self-disclosed investment experience and investment approaches, respectively. The difference test is the p-value associated with the t-test for differences in the coefficients of Disagreement between Professional and Novice in model (2) and Momentum and Value investors in model (3), respectively. The sample consists of 956 firms with overall 52,888 firm-month observations for the sample period between 2013 - 2017. The regressions are estimated using time, firm, and industry fixed effects. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm and month levels.

	1	2	3
Disagreement $_{Overall}$	-0.061*** [0.0120]		
Investors	-0.212*** [0.0261]		
*Disagreement $_{Overall}$ * Investors*	-0.079*** [0.0096]		
Disagreement $_{Professional}$		-0.054*** [0.0070]	
Disagreement $_{Intermediate}$		-0.048*** [0.0087]	
Disagreement $_{Novice}$		-0.037*** [0.0073]	
Disagreement $_{Momentum}$			-0.062*** [0.0075]
Disagreement $_{Technical}$			-0.051*** [0.0092]
Disagreement $_{Fundamental}$			-0.044*** [0.0056]
Disagreement $_{Value}$			-0.027*** [0.0051]
Media Coverage	-0.152*** [0.0225]	-0.207*** [0.0232]	-0.193*** [0.0239]
Analyst Coverage	0.113*** [0.0149]	0.119*** [0.0148]	0.120*** [0.0148]
Leverage	-0.031 [0.0232]	-0.086*** [0.0238]	-0.085*** [0.0244]
Adv/Sales	-0.013* [0.0071]	-0.013* [0.0068]	-0.013* [0.0070]
Market/Book Ratio	-0.040*** [0.0107]	-0.035*** [0.0112]	-0.033** [0.0114]
Firm Size	0.184*** [0.0566]	0.198*** [0.0579]	0.194*** [0.0585]
ROA	0.054*** [0.0150]	0.053*** [0.0160]	0.054*** [0.0159]
Earnings Volatility	0.014** [0.0050]	0.015** [0.0058]	0.015** [0.0060]
Sales Growth	0.077** [0.0327]	0.052 [0.0303]	0.053 [0.0305]
Real GDP $_{t-1}$	-0.070 [0.0399]	-0.076* [0.0408]	-0.076* [0.0406]
Diff (p-value)		(0.1347)	(0.003)***
Fixed Effects	Y	Y	Y
Adj. R-squared	0.247	0.232	0.235
Firms	956	956	956
Observations	52,888	52,888	52,888

Table 9: Disagreement and alternative proxies of Return Synchronicity

Notes: The table reports the regression results of the Disagreement and alternative proxies of Return Synchronicity. In Model (1) Return Synchronicity is calculated using the Carhart (1997) four-factor model and using the Adjusted R-squared. In Model (2) and (3) we use Morck et al. (2000) approach to calculate Return Synchronicity based on R-squared and adjusted R-squared. In Model (4) we follow Peng and Xion (2006) approach to calculate Return Synchronicity as a time series of Pearson correlation coefficients. All right-hand side variables are standardised. The sample consists of 956 firms, whereas firm-month observations vary based on the difference of approaches. The sample period is 2013 – 2017. The regressions are estimated using time, firm, and industry fixed effects. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm and month levels.

	1	**2**	**3**	**4**
	FF4-Adj	**MYY**	**MYY-Adj**	**CORR**
Disagreement	-0.059***	-0.027**	-0.026*	-0.024***
	[0.0122]	[0.0095]	[0.0119]	[0.0030]
Media Coverage	-0.241***	-0.170***	-0.187***	-0.054***
	[0.0291]	[0.0199]	[0.0223]	[0.0054]
Analyst Coverage	0.096***	0.089***	0.078***	0.031***
	[0.0149]	[0.0138]	[0.0158]	[0.0041]
Leverage	-0.127***	-0.074**	-0.088**	-0.029***
	[0.0283]	[0.0276]	[0.0346]	[0.0060]
Adv/Sales	-0.019***	-0.002	-0.007	-0.004***
	[0.0060]	[0.0058]	[0.0046]	[0.0008]
Market/Book Ratio	-0.028	-0.030**	-0.039**	-0.005
	[0.0162]	[0.0115]	[0.0132]	[0.0042]
Firm Size	0.217***	0.165**	0.167**	0.056***
	[0.0689]	[0.0544]	[0.0662]	[0.0131]
ROA	0.048**	0.045**	0.056*	0.010**
	[0.0216]	[0.0199]	[0.0257]	[0.0040]
Earnings Volatility	0.015	0.020*	0.027***	0.002
	[0.0114]	[0.0101]	[0.0085]	[0.0016]
Sales Growth	0.051	0.069**	0.087**	0.016*
	[0.0345]	[0.0271]	[0.0358]	[0.0077]
Real GDP $_{t-1}$	-0.083	-0.053	-0.058	-0.023*
	[0.0503]	[0.0369]	[0.0431]	[0.0117]
Fixed Effects	Y	Y	Y	Y
Adj. R-squared	0.168	0.342	0.286	0.270
Firms	956	956	956	956
Observations	44,548	52,686	47,266	52,916

Table 10: Return Synchronicity and Disagreement based on alternative prediction models

Notes: The table reports the regression results of the Disagreement and Return Synchronicity. The dependent variable is Return Synchronicity, which is calculated using the Carhart (1997) four-factor model using R-squared, adjusted R-squared and time series of Pearson correlation coefficient, respectively. Disagreement is derived from the users' recommendations predicted using Support Vector Machine and Maximum Entropy models. All right-hand side variables are standardised. The sample consists of 956 firms with variable firm-month observations in the Model (1) – (6) for the sample period between 2013 - 2017. The regressions are estimated using time, firm, and industry fixed effects. Variable definitions are presented in appendix B. *, **, and *** represent significance at 10%, 5%, and 1%, respectively. The standard errors are reported in brackets and are robust to heteroscedasticity and clustered at the firm and month levels.

	Support Vector Machine			**Maximum Entropy**		
	1	**2**	**3**	**4**	**5**	**6**
	FF4	**FF4-Adj**	**CORR**	**FF4**	**FF4-Adj**	**CORR**
Disagreement	-0.043**	-0.050***	-0.022***	-0.042***	-0.049***	-0.022***
	[0.015]	[0.015]	[0.004]	[0.013]	[0.013]	[0.003]
Media Coverage	-0.228***	-0.240***	-0.053***	-0.230***	-0.242***	-0.054***
	[0.023]	[0.028]	[0.005]	[0.024]	[0.029]	[0.005]
Analyst Coverage	0.116***	0.096***	0.031***	0.116***	0.095***	0.031***
	[0.015]	[0.015]	[0.004]	[0.014]	[0.015]	[0.004]
Leverage	-0.095***	-0.126***	-0.028***	-0.095***	-0.125***	-0.028***
	[0.024]	[0.028]	[0.006]	[0.024]	[0.028]	[0.006]
Adv/Sales	-0.015*	-0.019**	-0.004***	-0.015*	-0.019**	-0.004***
	[0.007]	[0.006]	[0.001]	[0.007]	[0.006]	[0.001]
Market/Book Ratio	-0.032**	-0.029*	-0.006	-0.032**	-0.028	-0.005
	[0.011]	[0.016]	[0.004]	[0.011]	[0.016]	[0.004]
Firm Size	0.200***	0.213**	0.054***	0.202***	0.216***	0.055***
	[0.058]	[0.069]	[0.013]	[0.058]	[0.069]	[0.013]
ROA	0.054***	0.048**	0.010**	0.054***	0.049**	0.010**
	[0.016]	[0.022]	[0.004]	[0.016]	[0.022]	[0.004]
Earnings Volatility	0.016**	0.015	0.002	0.016**	0.015	0.003
	[0.006]	[0.011]	[0.002]	[0.006]	[0.011]	[0.002]
Sales Growth	0.051	0.051	0.016*	0.050	0.049	0.015*
	[0.030]	[0.034]	[0.008]	[0.030]	[0.035]	[0.008]
Real GDP $_{t-1}$	-0.076*	-0.083	-0.023*	-0.076*	-0.084	-0.023*
	[0.041]	[0.050]	[0.012]	[0.041]	[0.050]	[0.012]
Fixed Effects	Y	Y	Y	Y	Y	Y
Adj. R-squared	0.227	0.168	0.269	0.227	0.168	0.269
Firms	956	956	956	956	956	956
Observations	52,888	44,548	52,916	52,888	44,548	52,916

Appendix A

StockTwits Recommendation Predictions and Text Analysis

For recommendation predictions of StockTwits ideas, we use the Random Forest Decision Trees model (RFDT). For RFDT to work as an ensemble, decision trees are created based on the impurity criterion. In our case, we use Entropy as the impurity criterion.

$$Entropy = -\sum_{i=1}^{L} f_i \log(f_i)\#$$

(1)

In equation (1), f_i is the frequency of label *i* at a node and *L* is the total number of unique labels. After calculating Entropy, the next step is to measure information gain, which is the metric that measures the expected reduction in the level of impurity in a given data set.

$$IG_{j,k} = Entropy_j - Entropy_{j,k}$$

(2)

In equation (2), $IG_{j,k}$ is the information gain from the given sample, *j* is the target value, *k* is split features, $Entropy_j$ is Entropy calculated for the target value and $Entropy_{j,k}$ is Entropy after the data are split based on k features. However, it is pertinent to note that even if the impurity criterion is changed to Gini impurity, our results remain constant, and there is no change in prediction outcome scores, such as AUC measures or F1 scores. To start applying the RFDT model, for each decision tree, the importance of the node is calculated based on the impurity criterion, as follows:

$$\theta_j = W_j C_j - W_j^{left} C_j^{left} - W_j^{Right} C_j^{Right}$$

(3)

Where θ_j is the importance of node *j*, W_j is the weighted number of samples reaching node *j*, C_j is the impurity value of node *j*. Similarly, $W_j^{left} C_j^{left}$ presents the values from the left node and $W_j^{Right} C_j^{Right}$ presents the values from the right node. In the next step, the importance of each feature *i* on the decision tree is calculated as follows:

$$\pi_i = \frac{\sum_{j:node\ j\ split\ on\ feature\ i} \theta_j}{\sum_{a\ \epsilon\ all\ nodes} \theta_a}$$

(4)

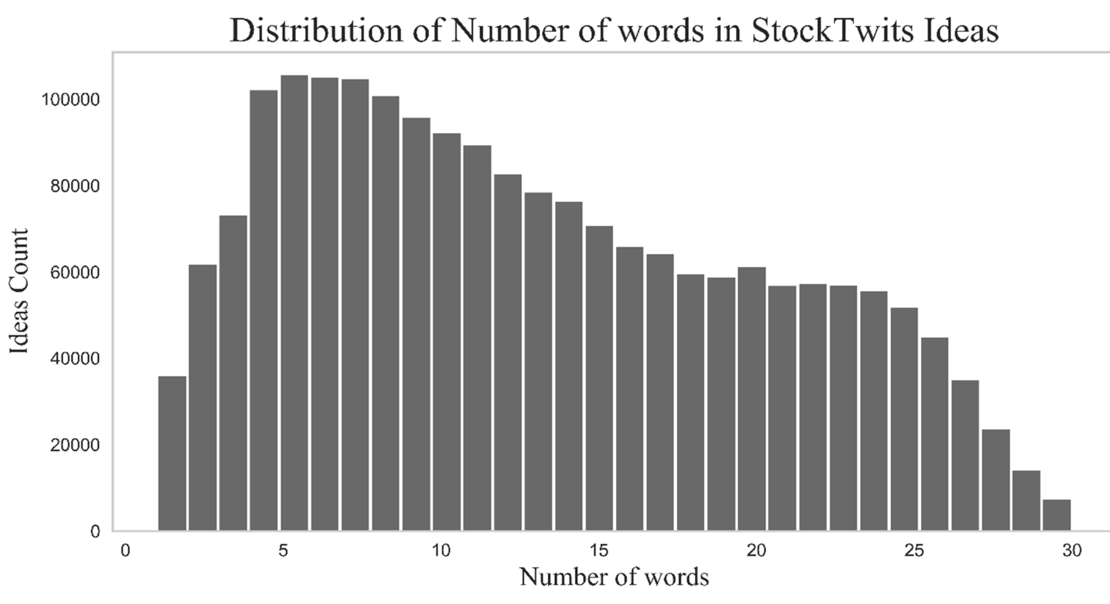
Where π_i is the importance of feature i, θ_j is the importance of node j from equation (3), and θ_a is the importance of all the nodes in a given tree. These values are then normalised, ranging between 0 and 1. Finally, overall feature importance i in all trees γ at the random forest level is calculated as an average of the features of all the decision trees.

$$\lambda_i = \frac{\sum_{\gamma \in all\ trees} \pi_i}{\sum DT} \tag{5}$$

In equation (5), λ_i is the importance of feature *i* calculated from all trees that are part of the random forest model, π_i is the normalized value of feature importance calculated in equation (4), and *DT* is the total number of decision trees that are part of the random forest ensembling process. Figure F2 in the internet appendix presents a sample decision tree.

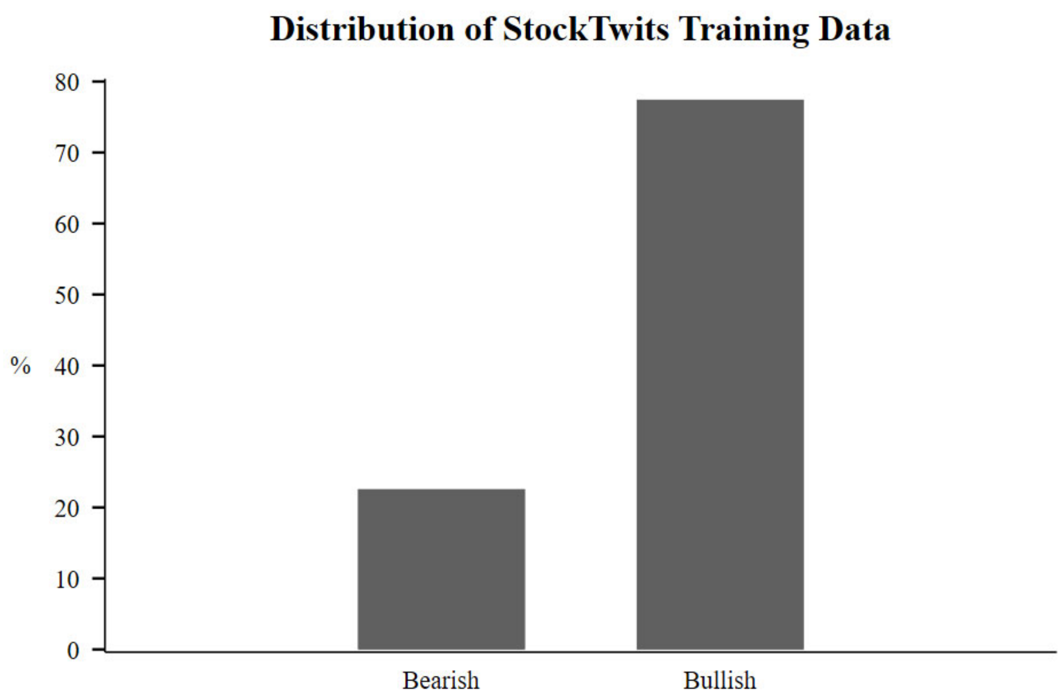
Text Analysis and Tokenisation

We use Baziotis, Pelekis, and Doulkeridis (2017) *Ekphrasis* library for text analysis. *Ekphrasis* is trained on Wikipedia and more than 330 million tweets. *Ekphrasis* also translates regular expressions used in StockTwits ideas in the form of emojis by using its social tokenizer. Table A3 in the internet appendix presents randomly selected ideas along with the bag of words. To further explore the quality of our StockTwits dataset, we plot the distribution of number of words in StockTwits ideas. Overall, the average number of words in the distribution is 13.01, with a median of 12 words. The standard deviation of the distribution is 7.55 words, whereas the minimum length of ideas is 1 and the maximum length of ideas is 115 words.



Training Prediction Models

We train our prediction model based on self-labelled[1] ideas from investors on StockTwits. The size of the training data set plays a vital role in the accurate prediction of investors' recommendations. Unlike Antweiler and Frank (2004) who used 1,000 messages for training, Giannini, Irvine, and Shu (2019) who used 2,000 Twitter posts for training and Cookson and Niessner (2019) who used 472,368 StockTwits ideas, the training data set used in our prediction models is 1.92 million StockTwits ideas. The distribution of our training data set is presented in the following figure.



Prediction Accuracy

In the RFDT model, our F1 score is 89%, and the AUC[2] score is 79%. Considering the importance of k-fold cross-validation, we test the prediction accuracy of our overall model using 10-fold cross-validation (CV). Our cross-validation accuracy is 82% and falls well within the range of robust prediction models. We select the best model based on the F1 score and CV accuracy. In addition to RFDT model, we also use Support Vector Machine (SVM) and

[1] This is important to control subjective bias when a data set is hand-labelled and then calibrated in prediction models.

[2] Area Under the Curve

Maximum Entropy (MaxEnt) to predict investors' recommendations. We use these models to check the robust of our results in section 6 of this study. The table below presents the results from three prediction models we use in this study.

Recommendation Prediction Models

Prediction Models	AUC	Accuracy	F1 Score
Random Forest Decision Trees	0.79	0.82	0.89
Support Vector Machine	0.81	0.80	0.87
Maximum Entropy	0.77	0.74	0.82

Following table presents the sample list of bullish and bearish words in the investors' recommendations on StockTwits after the sentiment analysis based on RFDT model.

Bullish and Bearish words after the Sentiment Analysis

Bullish		Bearish	
Aggressive	Thestreet	Announcement	Shoulders
Beat	Trade	Attention	Sinking
Earnings	Trades	Away	Spike
Forecast	Trailing	Buyback	Split
Fundamentals	Trending	Confidence	Spot
Gains	Tripled	Consolidation	Spread
Historical	Undervalued	Credit	Strategy
Hold	Upgrade	Decrease	Technology
Lawsuits	Uptrend	Disappointed	Topped
Leadership	Value	Downgraded	Trouble
Legal	Volatility	Downward	Tumble
Lock	Voting	Freefall	Understand
Long	Warren	Future	Unfortunately
Patent	Watching	Information	Unload
Psychological	Wave	Loss	Upcoming
Published	Wednesday	News	Vertical
Ratings	Weekend	Outlook	Volatility
Ratio	Welldone	Pullback	Wait
Research	Winner	Put	Wiped
Revisions	Wires	Resistance	Wish
Signal	Witnessing	Risky	Worry
Special	Worth	Rumors	Worst
Stochastic	Yield	Run	Worthless
Surprise		Sell	Wrong
Technology		Short	

Appendix B: Variables Definitions

Variables	Data Source	Frequency	Description
			StockTwits Variables
Disagreement	StockTwits	Monthly	Disagreement is calculated by predicting the investors' recommendations using the Random Forest Decision Trees analysis and then following Cookson and Niessner (2019). The level of disagreement ranges between 0 as agreement and 1 as complete disagreement between the investors on StockTwits.
Ideas			Natural logarithm of the number of ideas posted by the investors on StockTwits while discussing the sample firms.
Investors			Natural logarithm of the number of distinct users posting on StockTwits while discussing the sample firm.
Proximity			Natural logarithm of the number of distinct users from the same US State where the firm headquarter is located, posting on StockTwits while discussing the sample firm.
Network			The sum of distinct investors' followers, who post ideas while discussing the sample firms.
Popularity			Natural logarithm of the number of likes a user post receives while discussing the sample firm.
Discussion			Natural logarithm of the number of replies a post receives while discussing the sample firm.
$Revisions_t$			Natural logarithm of the sum of the number of times a distinct investor revise her recommendations (Bullish-to-Bearish) or vice versa on daily basis and then aggregated at monthly frequency.
$Revisions_{t-1}$			Natural logarithm of the lagged value of the sum of the number of times a distinct investor revise her recommendations (Bullish-to-Bearish) or vice versa on daily basis and then aggregated at monthly frequency
			Media Coverage
Breaking News	Ravenpack	Monthly	Natural logarithm of number of breaking news covered in the media related to the firm.
Full Articles			Natural logarithm of number of detailed articles and editorials published in the media while discussing the firm.
Press Release			Natural logarithm of number of press releases issued by the firm.
			Firm Level Characteristics
Return Synchronicity	CRSP	Monthly	Return synchronicity is calculated based on the value of the coefficient of determination (R^2). To derive the value of the coefficient of determination, we use Carhart (1997) four-factor model and take the log transformation of the coefficient of determination.
Firm Size		Quarterly	Natural logarithm of Market Capitalisation of the Firm.
Analyst Coverage	I/B/E/S	Monthly	Natural logarithm of Number of Analysts covering the firm.

Appendix B: Variables Definitions

Variables	Data Source	Frequency	Description
Leverage		Quarterly	Total long-term debt scaled by total assets.
Adv/Sales		Quarterly	Advertising to sales ratio of the sample firms.
Market/Book Ratio		Quarterly	Market value of equity scaled by the book value of the total assets.
Sales Growth		Quarterly	Natural logarithm of firm sales.
ROA		Quarterly	Income before extraordinary items scaled by total assets.
Earnings Volatility	Compustat	Quarterly	Standard Deviation of the ratio between income before extraordinary items in the current quarter and total assets in the previous quarter.
Firm Opacity		Yearly	Discretionary accruals used as a proxy of firm opacity or measure of firms' accrual quality calculated based on extended Jones (1991) model.
Diversity		Yearly	The natural logarithm of the number of business segments of the firm multiplied by the number of geographic segments.
Competition		Yearly	Industry competition based on firm level assets calculated using the Herfindahl-Hirschman Index.
Insider Trading	Thomson Reuters	Monthly	Insider trading is calculated as the difference between the buy and sell insider trades scaled by the total insider trades in a given month of a sample firm.
Real GDP	US Bureau of Economic Analysis	Monthly	Real Gross Domestic Product (GDP) of the United States.